



# Méthodes de type Galerkin discontinu pour la résolution numérique des équations de Maxwell en régime fréquentiel

Victorita Dolean, Hugo Fol, Stephane Lanteri, Serge Piperno

## ► To cite this version:

Victorita Dolean, Hugo Fol, Stephane Lanteri, Serge Piperno. Méthodes de type Galerkin discontinu pour la résolution numérique des équations de Maxwell en régime fréquentiel. [Rapport de recherche] RR-5904, INRIA. 2006. inria-00071362

**HAL Id: inria-00071362**

**<https://inria.hal.science/inria-00071362>**

Submitted on 23 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

*Méthodes de type Galerkin discontinu  
pour la résolution numérique  
des équations de Maxwell en régime fréquentiel*

Victorita Dolean — Hugo Fol — Stéphane Lanteri — Serge Piperno

**N° 5904**

Mai 2006

\_\_\_\_\_ Thème NUM \_\_\_\_\_

A large blue rectangle occupies the lower half of the page. Overlaid on it is a large, light gray stylized 'R' logo. To the right of the 'R', the words 'Rapport de recherche' are written in a white serif font. A horizontal gray brushstroke is positioned below the text.

*Rapport  
de recherche*





## Méthodes de type Galerkin discontinu pour la résolution numérique des équations de Maxwell en régime fréquentiel

Victorita Dolean<sup>\* †</sup>, Hugo Fol<sup>†</sup>, Stéphane Lanteri<sup>†</sup>, Serge Piperno<sup>‡†</sup>

Thème NUM — Systèmes numériques  
Projets Caiman

Rapport de recherche n° 5904 — Mai 2006 — 94 pages

**Résumé :** On s'intéresse à la résolution numérique des équations de Maxwell tridimensionnelles en régime harmonique par des méthodes de type Galerkin Discontinu (GD) en maillages tétraédriques non-structurés où l'on approche les inconnues du problème par des fonctions constantes par morceaux (méthode GD d'ordre 0 ou GDP0) ou linéaires par morceaux (méthode GD d'ordre 1 ou GDP1). De nombreux travaux ces dernières années ont démontré l'intérêt des formulations GD pour la modélisation numérique de phénomènes de propagation d'ondes en milieux hétérogènes. Dans cette étude, on adapte au domaine fréquentiel (ou régime harmonique) des méthodes GD centrées précédemment développées pour les mêmes équations résolues en domaine temporel. Dans ce rapport, on traite essentiellement de la formulation des méthodes en question en 1D et 3D, et de leur analyse théorique (dispersion numérique et caractère bien posé du problème discret). La difficulté principale réside dans le fait que l'on traite d'un problème posé dans le domaine complexe pour lequel des résultats d'inversibilité sont difficiles à prouver en l'absence d'hypothèses supplémentaires. On conclut en présentant une série de résultats numériques préliminaires en 1D et 3D, ces derniers visant essentiellement à valider les méthodes GD proposées. La question de la résolution (itérative ou directe) des systèmes linéaires obtenus sera traitée dans un prochain rapport.

**Mots-clés :** électromagnétisme numérique, équations de Maxwell en domaine fréquentiel, méthode de type Galerkin Discontinu, méthode d'éléments finis, méthode de volumes finis, maillages non-structurés tétraédriques.

<sup>\*</sup> Laboratoire J.A. Dieudonné, Université de Nice/Sophia Antipolis

<sup>†</sup> INRIA Sophia Antipolis, projet Caiman

<sup>‡</sup> Cermics, ENPC

# Discontinuous Galerkin methods for the numerical resolution of the frequency-domain Maxwell's equations

**Abstract:** This report is concerned with the numerical resolution of the 3D time-harmonic Maxwell's equations using Discontinuous Galerkin methods on unstructured tetrahedral meshes where the unknowns of the problem are approximated using piecewise constant functions or piecewise linear functions. Several recent works have clearly demonstrated the benefits of DG methods for the numerical modeling of wave propagation problems in heterogeneous media. In this work we adapt to the harmonic regime centered DG methods previously developed for the time-domain Maxwell equations. In this report, we mainly discuss the formulation of these methods and the analysis of their properties in the 1D and 3D cases (numerical dispersion, well posedness of the discrete problem). From the theoretical viewpoint, the main difficulty stems from the fact that we have to study a problem defined in the complex domain for which invertibility results are difficult to prove without some additional hypotheses. We conclude with a series of 1D and 3D numerical results that aim at validating the proposed methods. Aspects related to the resolution (either direct or iterative) of the complex coefficients linear systems will be the subject of a forthcoming report.

**Key-words:** computational electromagnetism, time-harmonic Maxwell's equations, Discontinuous Galerkin method, finite element method, finite volume method, unstructured tetrahedral meshes.

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>7</b>
<b>2</b>	<b>Formulation du problème</b>	<b>10</b>
2.1	Équations de Maxwell en régime temporel . . . . .	10
2.2	Équations de Maxwell en régime harmonique . . . . .	13
2.3	Conditions aux limites . . . . .	13
2.4	Formulation et étude préliminaire du problème discret . . . . .	14
2.4.1	Méthode de volumes finis . . . . .	15
2.4.2	Méthode de type Galerkin discontinu . . . . .	17
<b>3</b>	<b>Équations de Maxwell 1D</b>	<b>21</b>
3.1	Méthode de volumes finis . . . . .	23
3.1.1	Formulation du schéma . . . . .	23
3.1.2	Diagonalisation de la matrice de flux . . . . .	24
3.1.3	Traitement des conditions aux limites . . . . .	25
3.1.4	Étude du problème discret . . . . .	29
3.2	Méthode de type Galerkin discontinu . . . . .	31
3.2.1	Formulation du schéma . . . . .	31
3.2.2	Traitement des conditions aux limites . . . . .	33
3.2.3	Formulation matricielle . . . . .	35
3.2.4	Formulation $P_1$ . . . . .	38
3.2.5	Étude du problème discret . . . . .	40
<b>4</b>	<b>Équations de Maxwell 3D</b>	<b>41</b>
4.1	Méthode de volumes finis . . . . .	43
4.1.1	Traitement des conditions aux limites . . . . .	44
4.1.2	Formulation matricielle . . . . .	47
4.1.3	Étude de la dispersion numérique . . . . .	50
4.2	Méthode de type Galerkin discontinu . . . . .	53
4.2.1	Traitement des conditions aux limites . . . . .	55
4.3	Étude du problème discret . . . . .	59
4.3.1	Formulation $P_1$ . . . . .	59
4.3.2	Étude de la dispersion numérique . . . . .	60
<b>5</b>	<b>Résultats numériques</b>	<b>64</b>
5.1	Résultats 1D . . . . .	64
5.1.1	Maillage uniforme, milieu homogène . . . . .	65
5.1.2	Maillage non uniforme, milieu homogène . . . . .	67
5.1.3	Maillage uniforme, milieu hétérogène . . . . .	68
5.1.4	Maillage non uniforme, milieu hétérogène . . . . .	68
5.2	Résultats 3D . . . . .	72

<b>6</b>	<b>Conclusions et perspectives</b>	<b>72</b>
<b>7</b>	<b>Annexes</b>	<b>74</b>
7.1	Équivalences des conditions aux limites 1D . . . . .	74
7.2	Propriétés de la matrice $N_{jk}$ . . . . .	78
7.3	Décomposition de la matrice des flux . . . . .	79
7.4	Équivalences des conditions aux limites 3D . . . . .	81
7.5	Quelques propriétés des matrices antisymétriques . . . . .	83
7.6	Blocs matriciels du schéma GDP0 . . . . .	85
7.7	Blocs matriciels du schéma GDP1 . . . . .	86
7.7.1	Procédé de vectorisation . . . . .	86
7.7.2	Propriétés du produit de Kronecker . . . . .	87
7.7.3	Matrice 3D des gradients des fonctions de base P1 . . . . .	87
7.7.4	Propriétés matricielles . . . . .	89

## Notations

$\mathbb{K}$  représente le corps des réels  $\mathbb{R}$  ou celui des complexes  $\mathbb{C}$ .

Pour  $z \in \mathbb{C}$ ,  $\bar{z}$  représente son complexe conjugué.

Pour  $\mathbf{Z} \in \mathbb{C}^n$ , où  $\mathbf{Z} = (z_j)_{j=1,\dots,n}$ ,  $\mathbf{Z}^* = (\bar{z}_j)_{j=1,\dots,n}$ .

$|\cdot|$  désigne simultanément la norme euclidienne d'un élément de  $\mathbb{K}^{m \times n}$  et le cardinal d'un ensemble fini.

$\mathbf{M}_n(\mathbb{K})$  est l'espace vectoriel des matrices carrées sur  $\mathbb{K}$  de taille  $n \times n$ .

$0_n$  est la matrice nulle de l'espace  $\mathbf{M}_n(\mathbb{K})$ .

$\mathbf{I}_n$  est la matrice identité de l'espace  $\mathbf{M}_n(\mathbb{K})$ .

$\mathbf{D}_n(\mathbb{K})$  est le sous-espace vectoriel des matrices diagonales de  $\mathbf{M}_n(\mathbb{K})$ .

$\mathbf{GL}_n(\mathbb{K})$  est l'ensemble des matrices inversibles de  $\mathbf{M}_n(\mathbb{K})$ .

$\mathbf{M}_{n,m}(\mathbb{K})$  est l'espace vectoriel des matrices rectangulaires sur  $\mathbb{K}$  de taille  $n \times m$ .

$\mathbf{M}_{m \times n}(\mathbb{K})$  est l'espace vectoriel des matrices carrées sur  $\mathbb{K}$  de  $n \times n$  blocs de taille  $m$ .

$\mathbf{D}_{m \times n}(\mathbb{K})$  est le sous-espace vectoriel des matrices diagonales par blocs de  $\mathbf{M}_{m \times n}(\mathbb{K})$ .

Pour  $M = (m_{jk})_{\substack{1 \leq j \leq n \\ 1 \leq k \leq m}} \in \mathbf{M}_{n,m}(\mathbb{C})$ ,  $M^* = (\bar{m}_{kj})_{\substack{1 \leq j \leq n \\ 1 \leq k \leq m}}$  est la matrice adjointe de  $M$ . Si

$M \in \mathbf{M}_{n,m}(\mathbb{R})$ , alors  $M^* = M^t$ .

$\forall (X, Y) \in \mathbb{K}^{2(m \times n)} \cup \mathbb{K}^{2(mn)}$ ,  $X^*Y = \sum_{j=1}^{mn} \bar{x}_j y_j$  est le produit scalaire dans  $\mathbb{K}^{m \times n}$  et  $\mathbb{K}^{mn}$ .

$V_j$  est le volume occupé par la cellule  $C_j$ .

$dx$  et  $d\sigma$  sont respectivement les mesures de volume et de surface induite dans  $\mathbb{R}^3$ .

$\mathbf{n}_j \in \mathbb{R}^3$  est la normale unitaire sortante à la surface  $\partial C_j$  de  $C_j$ .

$\mathcal{I}$  est l'ensemble des indices des cellules purement internes à  $\Omega_h$ .

$\mathcal{I}_b$  est l'ensemble des indices des cellules dont au moins un des bords s'appuie sur  $\Gamma_h$ .

$\mathcal{V}_j^d = \{k \in \mathcal{I} / \Sigma_{jk} \subset \Omega_h \text{ pour } C_k \text{ voisine de } C_j\}$ .

$\mathcal{V}_j^a = \{k \in \mathcal{I}_b / \Sigma_{jk} \subset \Gamma_h^a \text{ pour } C_k \text{ voisine fictive de } C_j\}$ .

$\mathcal{V}_j^m = \{k \in \mathcal{I}_b / \Sigma_{jk} \subset \Gamma_h^m \text{ pour } C_k \text{ voisine fictive de } C_j\}$ .

$\mathcal{V}_j^b = \{k \in \mathcal{I}_b / \Sigma_{jk} \subset (\Gamma_h^m \cup \Gamma_h^a) \text{ pour } C_k \text{ voisine fictive de } C_j\}$ .

$\mathcal{V}_j = \mathcal{V}_j^d \cup \mathcal{V}_j^m \cup \mathcal{V}_j^a$ .

$\mathcal{V}_j^d \cap \mathcal{V}_j^m = \mathcal{V}_j^d \cap \mathcal{V}_j^a = \mathcal{V}_j^a \cap \mathcal{V}_j^m = \emptyset$ .

$\mathcal{I}_m = \{j \in \mathcal{I} / \mathcal{V}_j^m \neq \emptyset\}$  est l'ensemble des indices des cellules qui possèdent au moins un côté appartenant à  $\Gamma_h^m$ .

$\mathcal{I}_a = \{j \in \mathcal{I} / \mathcal{V}_j^a \neq \emptyset\}$  est l'ensemble des indices des cellules qui possèdent au moins un côté appartenant à  $\Gamma_h^a$ .

$\mathcal{V}_j = \{k \in \mathcal{I} \cup \mathcal{I}_b / C_k \text{ est voisine de } C_j \text{ et } \partial C_j \cap \partial C_k \text{ est une surface de mesure non nulle}\}$  est l'ensemble des cellules voisines de  $C_j$ .



$$\Sigma_{jk} = \partial C_j \cap \partial C_k, \quad k \in \mathcal{V}_j.$$

$$\partial C_j = \bigcup_{k \in \mathcal{V}_j} \Sigma_{jk}.$$

$S_{jk}$  est l'aire de la surface  $\Sigma_{jk}$ .

$\mathbf{n}_{jk} = (n_{jk}^{x_1}, n_{jk}^{x_2}, n_{jk}^{x_3})^t \in \mathbb{R}^3$  est la normale unitaire sortante à  $\Sigma_{jk}$  allant de  $C_j$  vers  $C_k$ .

$\forall (j, k) \in \mathcal{I} \times (\mathcal{I} \cup \mathcal{I}_b)$ ,  $\mathbf{n}_{jk} = -\mathbf{n}_{kj}$ .

## 1 Introduction

Les équations qui modélisent les problèmes de propagation d'ondes électromagnétiques c'est-à-dire, les équations de Maxwell, suscitent un engouement très vif tant de la part des physiciens que de la part des mathématiciens, et cela en raison de leurs nombreuses applications industrielles. En effet, les problèmes de furtivité radar, de compatibilité électromagnétique, de conception d'antennes ou de téléphonie cellulaire, pour n'en citer qu'un petit nombre, sont étudiés numériquement à partir des équations de Maxwell. La simulation numérique a contribué pour sa part à une meilleure compréhension de ces équations et c'est pour cette raison que de nombreuses recherches continuent dans cette voie afin de parvenir à des méthodes qui restituent au mieux leurs propriétés physiques et l'universalité de leur application. C'est là tout l'enjeu des méthodes numériques : restituer les propriétés physiques de ces équations et confronter ces méthodes à de nombreuses applications avec un souci de performance. On peut distinguer plusieurs classes de méthodes numériques suivant que l'on souhaite traiter les équations de Maxwell en domaine temporel en regardant l'évolution en temps du champ électromagnétique, ou les équations de Maxwell en domaine fréquentiel en regardant cette fois-ci le comportement du champ électromagnétique lorsque le terme source suit une dépendance harmonique en temps. Bien que ces deux formulations aient un lien physique étroit (voir par exemple [12]), les méthodes numériques développées pour leur approximation peuvent s'attaquer indistinctement aux deux systèmes d'équations ou au contraire être spécifiques à chacun. Les méthodes développées pour la résolution des équations en domaine temporel utilisent, le plus souvent, une formulation du premier ordre, alors que les méthodes pour la résolution des équations de Maxwell en domaine fréquentiel s'appuient habituellement sur la formulation du deuxième ordre dans laquelle on élimine le champ électrique ou le champ magnétique, afin de réduire la taille des systèmes algébriques résultant.

Indépendamment des formulations des équations de Maxwell, on peut distinguer trois grandes classes de méthodes numériques lorsqu'il s'agit de traiter de propagation d'ondes dans des milieux fortement hétérogènes : les méthodes de différences finies (DF), les méthodes d'éléments finis (EF), et enfin les méthodes de volumes finis (VF). La première méthode numérique efficace développée pour les équations de Maxwell a été une méthode DF due à Yee [39]. Cette méthode DFDT (différences finies dans le domaine temporel) conserve l'énergie, propriété que l'on cherche pour les méthodes numériques appliquées aux problèmes de propagation d'ondes en domaine temporel. De plus, sa grande simplicité d'implémentation la rend peu coûteuse en taille mémoire et en temps de calcul et on peut, par exemple pour le problème de la diffraction, effectuer un calcul pour une signal large bande, et étendre par transformation de Fourier le résultat à toutes les fréquences du spectre. Malheureusement, l'utilisation d'un maillage cartésien contraint le pas d'espace à être petit pour tenir compte de la géométrie des objets représentés dès lors que ceux-ci ont des formes complexes, ce qui réduit par la même occasion le pas de temps pour les schémas explicites comme le schéma de Yee (condition de stabilité de type CFL). D'autre part, la représentation en

maillages cartésiens non uniformes engendre des ondes parasites qui perturbent les solutions numériques obtenues.

L'autre méthode importante développée une quinzaine d'années plus tard fut la méthode des EF d'arêtes [27] qui fut déclinée en plusieurs versions [25, 28]. Cette méthode présente beaucoup d'avantages, avec entre autres la faculté de traiter des maillages non structurés et donc des géométries complexes, de conserver la propriété de divergence nulle au niveau discret, et de conserver l'énergie. Elle peut même être utilisée en tant que méthode d'ordre élevé [8, 20, 25]. Un point faible majeur réside dans l'inversion de la matrice de masse obtenue à chaque pas de temps en domaine temporel, ce qui amène à envisager des techniques de condensation de masse [7]. Cet inconvénient a également conduit à l'étude des schémas de type VF pour les problèmes de propagation d'ondes en domaine temporel en raison de leur caractère hyperbolique, en s'inspirant des méthodes développées pour les problèmes issus de la mécanique des fluides. Tout comme les EF, les méthodes VF ont l'avantage de pouvoir s'adapter facilement à des maillages non structurés. Leur particularité réside dans la formulation du flux numérique, qui peut être décentré, ce qui pour les équations de Maxwell rend le schéma diffusif [31] et peu dispersif, ou centré mais dans ce cas le schéma est non diffusif mais davantage dispersif [35].

Une classe de méthodes numériques relativement récente, appelée méthodes de type Galerkin Discontinu (GD), combine les avantages des méthodes d'éléments finis (EF) et de volumes finis (VF) puisqu'elle traite la discontinuité entre cellules voisines par une approximation VF sur les flux et qu'elle approche le champ dans chaque cellule par une base locale de fonctions. Cela permet une représentation plus riche des solutions recherchées, mais les méthodes GD sont à ce titre plus coûteuses que les méthodes EF standards dans la mesure où une inconnue du problème repose sur plusieurs degrés de liberté locaux. Dans ce formalisme, la méthode GD d'ordre 0 est considérée comme une méthode de type VF étant donné que les fonctions de base utilisées sont constantes. Ces méthodes ont été proposées initialement pour résoudre l'équation scalaire du transport des neutrons [19]. Dans le domaine temporel, ces méthodes permettent d'approcher efficacement plusieurs classes de problèmes hyperboliques lorsqu'on les combine à une intégration temporelle de type Runge-Kutta [6].

Dans le domaine de la propagation d'ondes, et plus particulièrement pour l'équation de Helmholtz, les méthodes EF ne peuvent éviter un effet de pollution numérique lorsque le nombre d'onde augmente [4]-[11] et ce quel que soit le type des fonctions de base utilisées. A cet effet, les méthodes GD ont montré une nette amélioration en prenant des ondes planes comme fonctions de base et en imposant une continuité faible à l'aide de multiplicateurs de Lagrange [9]. Pour ce qui concerne les équations de Maxwell en domaine temporel, de nombreux schémas reposent soit sur une formulation à flux centrés [10]-[32], soit sur une formulation à flux décentrés [16]. Elles ont même été utilisées avec une interpolation d'ordre très élevé (polynômes du dixième degré) pour approcher les équations de Maxwell en domaine temporel [15]. Le principal problème récurrent à ces travaux est que l'intégration en temps explicite contraint à prendre une condition CFL restrictive sur des maillages fins. Ce constat a conduit certains auteurs à développer des méthodes spatio-temporelles qui reposent sur l'utilisation d'un schéma semi-explicite pour les systèmes hyperboliques linéaires

en milieu inhomogène [26]. La condition CFL devient alors une CFL locale qui ne dépend plus que d'un petit nombre d'éléments, mais le décentrage du flux introduit malheureusement de la diffusion numérique. La solution est donc altérée après un nombre de pas de temps significatif.

Toujours dans le domaine de l'électromagnétisme mais cette fois-ci dans l'étude du problème aux valeurs propres discret, les méthodes GD ont aussi montré leur efficacité [17]. L'approximation discontinue et la nature non conforme de l'approximation permettent en l'occurrence de résoudre un problème aux valeurs propres standard puisque la matrice de masse est alors diagonale par blocs, plutôt que de résoudre un problème aux valeurs propres généralisé. Cependant, bien que des exemples 2D montrent que l'on n'obtient pas de valeurs propres parasites grâce à l'équivalence en 2D entre les solutions du problème aux valeurs propres discret pour les équations de Maxwell du premier ordre et les solutions de l'équation de Helmholtz, les exemples 3D ne peuvent éviter cet écueil à cause de la mauvaise représentation du noyau de l'opérateur rot-rot [18].

Les méthodes GD ont également connu un développement très important pour des équations de nature elliptique et parabolique [1]-[23]-[38] et cela indépendamment de leur développement pour les équations hyperboliques. Elles ont toutes la particularité de reposer sur l'inclusion d'un terme de pénalisation au sein du flux numérique sans lequel la convergence ne serait pas assurée. Toutes ces méthodes répertoriées et étudiées dans [2] ont des avantages et des inconvénients, mais restent supérieures aux méthodes EF classiques sur le plan de l'adaptativité *hp* et de leur utilisation sur des maillages non conformes. Un autre avantage est de pouvoir utiliser une base de fonctions nodales qui peuvent approcher des solutions dans des zones du domaine fortement singulières, contrairement aux éléments finis classiques [3]-[25].

Bien que les méthodes EF pour traiter les équations de Maxwell en domaine fréquentiel considèrent pour la majorité la formulation du second ordre [30]-[29]-[21], certaines études se basent sur la formulation du premier ordre comme dans [12] et [5]. Le passage des équations de Maxwell temporelles aux équations de Maxwell fréquentielles se fait grâce à une transformation de Laplace appliqué à un système perturbé, et en admettant le principe d'amplitude limite. Ce principe stipule que le champ électromagnétique du problème instationnaire est décomposé en la somme d'une onde harmonique liée à la pulsation de l'onde incidente, de plusieurs ondes harmoniques liées aux fréquences propres du problème, et enfin, d'une onde évanescence. Dans [12], le schéma utilisé est un schéma GD d'ordre 0, basé sur une fonction de flux numérique décentrée. C'est grâce à cette formulation du flux numérique, rendant coercive la forme bilinéaire associée, que des résultats d'existence et d'unicité de la solution discrète sur des domaines bornés et non-bornés peuvent être établis [13]-[14]. Des preuves de convergence ont même été établies lorsque la perturbation est nulle, mais cela nécessite de repasser par la formulation continue de manière à exploiter des résultats provenant de la théorie du contrôle.

L'objet de l'étude qui nous intéresse ici est d'appliquer aux équations de Maxwell en domaine fréquentiel des méthodes GD basées sur une fonction de flux numérique centrée, précédemment mises au point pour la résolution numérique des équations de Maxwell en

domaine temporel[35]-[33]-[32]. C'est une des raisons pour lesquelles nous utiliserons une formulation du premier ordre des équations de Maxwell. La suite de ce rapport est organisée en quatre sections. La section 2 précise le cadre mathématique de cette étude. On commence par rappeler les équations de Maxwell en domaine temporel avant d'introduire les équations de Maxwell en domaine fréquentiel via le principe d'amplitude limite. On définit ensuite le problème au limites associé aux équations de Maxwell en domaine fréquentiel et on termine cette section par une analyse préliminaire des problèmes discrets qui caractérisent les méthodes GD d'ordre 0 (GDP0 ou méthode de volumes finis) et d'ordre 1 (GDP1). La section 3 est consacrée à l'étude des formulations GDP0 et GDP1 appliquées aux équations de Maxwell 1D. On étudie notamment l'inversibilité des systèmes matriciels résultant des formulations GDP0 et GDP1 en 1D. La section 4 traite des formulations GDP0 et GDP1 pour les équations de Maxwell 3D. Dans cette section, on analyse la dispersion des ces formulations. Des résultats numériques préliminaires en 1D et 3D sont présentés dans la section 5. Enfin, la section 6 conclut cette étude.

## 2 Formulation du problème

### 2.1 Équations de Maxwell en régime temporel

Soit  $\Omega$  un ouvert régulier borné de  $\mathbb{R}^3$  de frontière  $\Gamma = \Gamma^m \cup \Gamma^a$  (avec  $\Gamma^m \cap \Gamma^a = \emptyset$ ) comme représenté sur la figure 1.

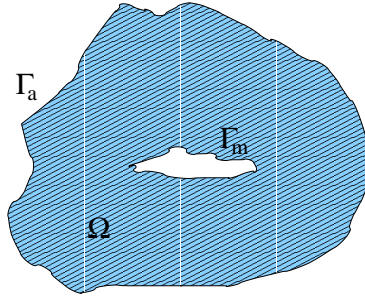


FIG. 1 – Domaine de calcul en 3D

Les lois de la physique, complétées par des relations phénoménologiques décrivant le comportement du milieu, fournissent un modèle pour la propagation des ondes électroma-

gnétiques dans ce milieu. Pour commencer, les lois d'Ampère, Faraday, ainsi que la loi de conservation de la charge s'écrivent comme suit :

$$\left\{ \begin{array}{lcl} \frac{\partial \mathbf{D}}{\partial t} & - & \text{rot} \mathbf{H} = -\mathbf{J}, \\ \frac{\partial \mathbf{B}}{\partial t} & + & \text{rot} \mathbf{E} = 0, \\ \frac{\partial \rho}{\partial t} & + & \text{div} \mathbf{J} = 0, \end{array} \right. \quad (1)$$

avec :

$$\left\{ \begin{array}{ll} \mathbf{E} = \mathbf{E}(\mathbf{x}, t) & \text{le champ électrique,} \\ \mathbf{D} = \mathbf{D}(\mathbf{x}, t) & \text{le déplacement électrique,} \\ \mathbf{H} = \mathbf{H}(\mathbf{x}, t) & \text{le champ magnétique,} \\ \mathbf{B} = \mathbf{B}(\mathbf{x}, t) & \text{l'induction magnétique,} \\ \mathbf{J} = \mathbf{J}(\mathbf{x}, t) & \text{la densité de courant électrique,} \\ \rho = \rho(\mathbf{x}, t) & \text{la densité de charge électrique.} \end{array} \right.$$

On a également que  $\text{div} \mathbf{B} = 0$  (ce qui exprime l'absence de charge magnétique libre) et  $\text{div} \mathbf{D} = \rho$ . Pour fermer le système (1) on a besoin de lois reliant  $\mathbf{D}$  et  $\mathbf{B}$  à  $\mathbf{E}$  et  $\mathbf{H}$  dépendant du matériau. Dans le cas de matériaux linéaires, on suppose généralement l'existence de matrices réelles  $\varepsilon(\mathbf{x}, t)$ ,  $\mu(\mathbf{x}, t)$  symétriques définies positives et  $\sigma(\mathbf{x}, t)$  réelle symétrique positive, appelées respectivement tenseur de permittivité électrique, de perméabilité magnétique et de conductivité électrique du milieu, qui relient les champs entre eux par :

$$\mathbf{D} = \varepsilon \mathbf{E}, \quad \mathbf{B} = \mu \mathbf{H} \quad \text{et} \quad \mathbf{J} = \sigma \mathbf{E}.$$

Dans ce qui suit, on considère le cas où  $\varepsilon(\mathbf{x})$ ,  $\mu(\mathbf{x})$  et  $\sigma(\mathbf{x})$  sont des variables scalaires (matériaux isotropes) et indépendantes du temps. Les quantités  $\varepsilon(\mathbf{x})$  et  $\mu(\mathbf{x})$  s'expriment en fonction des indices du vide  $\varepsilon_0$  et  $\mu_0$ , et des valeurs adimensionnées relatives au milieu considéré  $\varepsilon_r$  et  $\mu_r$ . On introduit enfin une constante appelée l'impédance du vide notée  $z_0$ . En résumé :

$$\left\{ \begin{array}{lcl} \varepsilon(\mathbf{x}) & = & \varepsilon_0 \varepsilon_r(\mathbf{x}), \\ \mu(\mathbf{x}) & = & \mu_0 \mu_r(\mathbf{x}), \\ c_0 & = & \frac{1}{\sqrt{\varepsilon_0 \mu_0}}, \quad c_r(\mathbf{x}) = \frac{1}{\sqrt{\varepsilon_r(\mathbf{x}) \mu_r(\mathbf{x})}} \quad \text{et} \quad c(\mathbf{x}) = c_0 c_r(\mathbf{x}), \\ z_0 & = & \sqrt{\frac{\mu_0}{\varepsilon_0}}, \quad z_r(\mathbf{x}) = \sqrt{\frac{\mu_r(\mathbf{x})}{\varepsilon_r(\mathbf{x})}} \quad \text{et} \quad z(\mathbf{x}) = z_0 z_r(\mathbf{x}). \end{array} \right. \quad (2)$$

Si on note :

$$N_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \quad N_y = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad N_z = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

et :

$$G_l = \begin{pmatrix} 0_3 & N_l \\ -N_l & 0_3 \end{pmatrix} \quad \text{avec } l \in \{x, y, z\},$$

alors les lois d'Ampère et Faraday peuvent se réécrire sous la forme matricielle suivante :

$$\partial_t(G_0 \mathbf{W}) + G_x \partial_x \mathbf{W} + G_y \partial_y \mathbf{W} + G_z \partial_z \mathbf{W} + K \mathbf{W} = 0, \quad (3)$$

où on a adopté la notation classique  $\partial_{\{x,y,z,t\}} A = \frac{\partial A}{\partial \{x,y,z,t\}}$  et où :

$$\mathbf{W} = \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}, \quad G_0 = \begin{pmatrix} \varepsilon(\mathbf{x}) & 0 \\ 0 & \mu(\mathbf{x}) \end{pmatrix} \quad \text{et} \quad K = \begin{pmatrix} \sigma(\mathbf{x}) & 0 \\ 0 & 0 \end{pmatrix}.$$

La solution de ces équations s'appelle *champ électromagnétique total*. L'intérêt de l'écriture des équations de Maxwell comme (3) est que cette forme est celle des systèmes hyperboliques symétriques du premier ordre ou systèmes de Friedrichs.

Bien que ces équations soient très générales, on se restreindra par la suite au problème de la *diffraction* par un obstacle : en plongeant un corps de caractéristiques données dans le vide en présence d'une onde incidente, il s'agit de déterminer la perturbation du *champ incident* induite par le corps, appelée *champ diffracté*. On a donc que  $\mathbf{W}^{\text{dif}} = \mathbf{W}^{\text{tot}} - \mathbf{W}^{\text{inc}}$  où  $\mathbf{W}^{\text{tot}}$  est le champ total en présence de l'obstacle et  $\mathbf{W}^{\text{inc}}$  est le champ incident en l'absence de celui-ci. On suppose que le champ incident vérifie l'équation (3) dans le vide si bien que le champ diffracté (qu'on notera abusivement  $\mathbf{W}$ ) vérifie l'équation (voir par exemple Helluy [12] pour plus de détails) :

$$\begin{aligned} \partial_t(G_0 \mathbf{W}) + G_x \partial_x \mathbf{W} + G_y \partial_y \mathbf{W} + G_z \partial_z \mathbf{W} + K \mathbf{W} &= -K \mathbf{W}^{\text{inc}} + (\mathbf{I}_6 - G_0) \partial_t \mathbf{W}^{\text{inc}} \\ &\equiv S(\mathbf{W}^{\text{inc}}). \end{aligned} \quad (4)$$

où  $\mathbf{I}_6$  désigne la matrice identité de l'espace  $\mathbf{M}_6(\mathbb{R})$ . Si on note par  $G_{\mathbf{n}} = n^x G_x + n^y G_y + n^z G_z$  où  $\mathbf{n}$  désigne un vecteur non-nul de  $\mathbb{R}^3$ , le fait que le système des équations de Maxwell (4) est hyperbolique signifie que la matrice de flux  $\bar{G}_{\mathbf{n}} = G_{\mathbf{n}} G_0^{-1}$  est diagonalisable dans  $\mathbb{R}^6$  et on a :

$$\begin{cases} \bar{G}_{\mathbf{n}} &= T_{\mathbf{n}} \Lambda_{\mathbf{n}} T_{\mathbf{n}}^{-1} \\ \bar{G}_{\mathbf{n}}^{\pm} &= T_{\mathbf{n}} \Lambda_{\mathbf{n}}^{\pm} T_{\mathbf{n}}^{-1} \\ \Lambda_{\mathbf{n}}^+ &= \text{diag}(\max(\lambda_{\mathbf{n},k}, 0)) \\ \Lambda_{\mathbf{n}}^- &= \text{diag}(\min(\lambda_{\mathbf{n},k}, 0)) \end{cases}$$

où  $\lambda_{\mathbf{n},k}$ ,  $k = 1, \dots, 6$  désignent les valeurs propres de la matrice de flux  $G_{\mathbf{n}}$ .

## 2.2 Équations de Maxwell en régime harmonique

Si le champ incident suit une dépendance harmonique en temps, c'est-à-dire s'il s'exprime sous la forme  $\mathbf{W}^{\text{inc}}(\mathbf{x}, t) = \mathbf{W}^{\text{inc}}(\mathbf{x})e^{i\omega t}$ , le comportement asymptotique en temps du champ diffracté sera déterminé en supposant le *principe d'amplitude limite* vrai (voir notamment Rauch [34] pour les limites d'applicabilité de ce principe). Ce qui veut dire que le champ diffracté est une superposition d'une onde harmonique d'amplitude  $\mathbf{W}(\mathbf{x})$  avec la même pulsation que le champ incident, d'ondes harmoniques correspondant aux fréquences propres  $\omega_k$  du problème (d'amplitudes  $\mathbf{W}_k$ ) et d'une onde évanescence. Si le principe d'amplitude limite est vrai, les  $\mathbf{W}_k$  sont à support compact et l'unique inconnue du problème asymptotique est  $\mathbf{W}$ . On rappelle dans ce qui suit des résultats énoncés dans Helluy [12], qui montrent le lien entre les équations en domaine temporel et les équations harmoniques.

**Théorème 1** *Principe d'amplitude limite.*

(i) Si on note  $p = i\omega + \nu$ ,  $\nu > 0$  et  $M_\nu(u) = \nu L_p(u)$  où  $L_p(u)$  est la transformée de Laplace partielle par rapport au temps au point  $p$ , et si :

$$\mathbf{W}(\mathbf{x}, t) = \mathbf{W}(\mathbf{x})e^{i\omega t} + \sum_{k=1}^m \mathbf{W}_k e^{i\omega_k t} + \phi(\mathbf{x}, t) \quad \text{tel que} \quad \lim_{t \rightarrow \infty} \phi(\mathbf{x}, t) = 0,$$

alors  $\lim_{\nu \rightarrow 0} M_\nu(\mathbf{W}(\mathbf{x}, t)) = \mathbf{W}(\mathbf{x})$  dans  $L^2(\Omega)$ .

(ii) Si  $\mathbf{W}(\mathbf{x}, t)$  est solution des équations (4) alors la quantité  $\mathbf{W}_\nu(\mathbf{x}) = M_\nu(\mathbf{W}(\mathbf{x}, t))$  vérifie les équations harmoniques perturbées :

$$pG_0\mathbf{W}_\nu + G_x\partial_x\mathbf{W}_\nu + G_y\partial_y\mathbf{W}_\nu + G_z\partial_z\mathbf{W}_\nu + K\mathbf{W}_\nu = S_\nu(\mathbf{W}^{\text{inc}}), \quad (5)$$

avec  $S_\nu(\mathbf{W}^{\text{inc}}) = -K\mathbf{W}^{\text{inc}} + (I_6 - G_0)p\mathbf{W}^{\text{inc}}$ .

La limite au sens des distributions (quand  $\nu$  tend vers zero) des équations (5) donne les *équations harmoniques* de solution  $\mathbf{W}(\mathbf{x})$  qui est l'amplitude de l'onde harmonique de pulsation  $\omega$  (principe de l'amplitude limite). Par la suite, si nécessaire (par exemple pour démontrer que le problème discret est bien posé) on étudiera plutôt le problème (5) en admettant les principes d'absorption limite et d'amplitude limite.

## 2.3 Conditions aux limites

**Condition réfléchissante.** On suppose que la partie  $\Gamma_m = \partial\Omega_m$  de la frontière du domaine de calcul et la surface d'un obstacle métallique. On y impose que la trace tangentielle du champ électrique est nulle c'est-à-dire  $\mathbf{E} \times \mathbf{n} = 0$  où  $\mathbf{n}$  désigne la normale à  $\Gamma_m$ .



**Condition d'impédance.** Il s'agit d'une condition de la forme suivante :

$$\mathbf{n} \times \mathbf{E} + \alpha \mathbf{n} \times \mathbf{n} \times \mathbf{H} = 0.$$

**Condition absorbante.** Sur la partie  $\Gamma_a = \partial\Omega_a$ , qui est en fait la frontière artificielle du domaine, on va imposer une condition absorbante, c'est-à-dire qu'il n'y a pas d'ondes entrantes dans le domaine. Puisque c'est la matrice  $\bar{G}_{\mathbf{n}} = G_{\mathbf{n}} G_0^{-1}$  qui est diagonalisable, cette condition s'écrit  $\bar{G}_{\mathbf{n}} G_0 \mathbf{W} = 0$ <sup>1</sup> On vérifie aisément (voir aussi l'annexe 7.3) que  $\bar{G}_{\mathbf{n}} = G_{\mathbf{n}}^- G_0^{-1}$  si bien que la condition à imposer s'écrit aussi  $G_{\mathbf{n}}^- \mathbf{W} = 0$  où  $\mathbf{n}$  désigne la normale sortante à  $\Gamma_a$ . On constate que celle-ci est équivalente à :

$$\mathbf{n} \times \mathbf{E} + \mathbf{n} \times \mathbf{n} \times \mathbf{H} = 0.$$

Il s'agit d'une condition de type impédance avec  $\alpha = 1$  (aussi appelée condition de type Silver-Müller d'ordre 1) qui est en fait une condition transparente approchée.

On considère par la suite le *problème de Maxwell harmonique* 3D posé sur le domaine  $\Omega$ , auquel on ajoute une condition réfléchissante sur  $\Gamma^m$  et une condition absorbante de type Silver-Müller d'ordre 1 sur  $\Gamma^a$ .

## 2.4 Formulation et étude préliminaire du problème discret

On considère maintenant l'équation perturbée (5) et on suppose  $\Gamma = \Gamma^a$ . On doit alors résoudre le problème aux limites suivant :

$$\begin{cases} pG_0 \mathbf{W} + G_x \partial_x \mathbf{W} + G_y \partial_y \mathbf{W} + G_z \partial_z \mathbf{W} + K \mathbf{W} = \mathbf{S} & \text{dans } \Omega, \\ G_{\mathbf{n}}^- \mathbf{W} = 0 & \text{sur } \partial\Omega, \end{cases} \quad (6)$$

où  $p = \nu + i\omega$  avec  $\nu > 0$ , qui peut aussi se mettre sous la forme :

$$\begin{cases} pG_0 \mathbf{W} + \operatorname{div} F(\mathbf{W}) + K \mathbf{W} = \mathbf{S} & \text{dans } \Omega, \\ G_{\mathbf{n}}^- \mathbf{W} = 0 & \text{sur } \partial\Omega. \end{cases} \quad (7)$$

avec  $F(\mathbf{W}) = (G_x \mathbf{W}, G_y \mathbf{W}, G_z \mathbf{W})^t$ .

On désigne par  $\Omega_h$  une approximation polyédrique du domaine  $\Omega$  et on suppose avoir découpé  $\Omega_h$  en une partition finie  $\mathcal{I} \subset \mathbb{N}$  de cellules volumiques  $C_j$  (pouvant être par exemple tétraédriques ou hexahédriques) :  $\bar{\Omega}_h = \bigcup_{j \in \mathcal{I}} C_j$ . Le maillage correspondant est désigné par  $\mathcal{T}_h$ . Tout comme pour  $\Omega$ ,  $\Omega_h$  possède des frontières de type métallique  $\Gamma_h^m$  et de type absorbante  $\Gamma_h^a$ , vérifiant  $\partial\Omega_h = \Gamma_h^m \cup \Gamma_h^a$  avec  $\Gamma_h^m \cap \Gamma_h^a = \emptyset$ .

<sup>1</sup>Le système des équations de Maxwell (4) est hyperbolique ce qui signifie que la diagonalisation s'opère sur la matrice de flux obtenue lorsque le vecteur d'état est  $\mathbf{Q} = G_0 \mathbf{W}$ . En d'autres termes, le système de référence pour la diagonalisation est le système (4) sous forme conservative :  $\partial_t(\mathbf{Q}) + \partial_x(\bar{G}_x \mathbf{Q}) + \partial_y(\bar{G}_y \mathbf{Q}) + \partial_z(\bar{G}_z \mathbf{Q}) + K G_0^{-1} \mathbf{Q} = \bar{S}(\mathbf{Q}^{\text{inc}})$ .

Dans chaque élément, la restriction de  $\mathbf{W}$  à la cellule  $C_j$  appartient à l'espace :

$$V_h = \{\mathbf{W} \in L^2(\Omega)^6 \mid \forall j, \mathbf{W}|_{C_j} \in \mathbb{P}_m(C_j)^6\}.$$

#### 2.4.1 Méthode de volumes finis

On considère ici le cas  $m = 0$  qui correspond à l'approximation volumes finis (cette méthode est étudiée plus en détail dans la section 2.4.1). L'approximation est donc discontinue, aucune relation de continuité n'étant imposée entre les éléments. On note  $\mathbf{W}_j$  la valeur moyenne du champ  $\mathbf{W}$  sur la cellule  $C_j$  :

$$\mathbf{W}_j = \frac{1}{V_j} \int_{C_j} \mathbf{W} dx.$$

Si on intègre l'équation (7) sur la cellule  $C_j$  on obtient :

$$\int_{C_j} (pG_0 + K) \mathbf{W} dx + \int_{C_j} \operatorname{div} F(\mathbf{W}) dx = \int_{C_j} \mathbf{S} dx,$$

ou bien encore :

$$\int_{C_j} (pG_0 + K) \mathbf{W} dx + \int_{\partial C_j} F(\mathbf{W}) \cdot \mathbf{n} ds = \int_{C_j} \mathbf{S} dx. \quad (8)$$

On utilise une fonction de flux numérique centrée pour les interfaces internes et décentrée pour les interfaces frontières absorbantes et on note par  $G_{jk} = \sum_{i=1}^3 n_{jk}^{x_i} G_i$  où  $\mathbf{n}_{jk} = (n_{jk}^{x_1}, n_{jk}^{x_2}, n_{jk}^{x_3})^t$  est la normale unitaire à l'interface entre les cellules  $C_j$  et  $C_k$ , dirigée de  $C_j$  vers  $C_k$ . On obtient alors :

$$\begin{aligned} (8) &\Leftrightarrow \int_{C_j} (pG_0 + K) \mathbf{W} dx + \sum_{j \in \mathcal{V}_j} \int_{\Sigma_{jk}} F(\mathbf{W}) \cdot \mathbf{n} ds = \int_{C_j} \mathbf{S} dx \\ &\Leftrightarrow \int_{C_j} (pG_0 + K) \mathbf{W} dx + \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} G_{jk} \left( \frac{\mathbf{W}_j + \mathbf{W}_k}{2} \right) ds \\ &\quad + \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} (G_{jk}^+ \mathbf{W}_j + G_{jk}^- \mathbf{W}_k) ds = \int_{C_j} \mathbf{S} dx. \end{aligned} \quad (9)$$

Comme nous le verrons plus loin, la condition absorbante exprimée faiblement correspond à  $G_{jk}^- \mathbf{W}_k = 0$ . Prenant cela en compte dans (9), on obtient :

$$(9) \Leftrightarrow \int_{C_j} (pG_0 + K) \mathbf{W} dx + \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} G_{jk} \left( \frac{\mathbf{W}_j + \mathbf{W}_k}{2} \right) ds + \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} G_{jk}^+ \mathbf{W}_j ds = \int_{C_j} \mathbf{S} dx. \quad (10)$$

On peut alors démontrer le lemme suivant.

**Lemme 1** *La solution du problème homogène associé à (7) vérifie la relation suivante :*

$$\frac{1}{2} \int_{\partial\Omega_h} (|\mathbf{Gn}| \mathbf{W}, \mathbf{W}) ds + \int_{\Omega_h} ((\nu G_0 + K) \mathbf{W}, \mathbf{W}) dx = 0, \quad (11)$$

et on peut conclure que l'unique solution du problème discret homogène est la solution nulle.

**Preuve** Comme  $\mathbf{W}$  est constant sur la cellule  $C_j$  en prenant le produit scalaire de l'équation homogène associée à (10) avec  $\mathbf{W}_j$  on obtient :

$$\int_{C_j} ((pG_0 + K) \mathbf{W}_j, \mathbf{W}_j) dx + \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} (G_{jk} \left( \frac{\mathbf{W}_j + \mathbf{W}_k}{2} \right), \mathbf{W}_j) ds + \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} (G_{jk}^+ \mathbf{W}_j, \mathbf{W}_j) ds = 0.$$

Si  $C_j$  ne s'appuie pas sur la frontière  $\partial\Omega_h$  la relation précédente s'écrit (en tenant compte du fait que  $\sum_k G_{jk} = 0$ ) :

$$\int_{C_j} ((pG_0 + K) \mathbf{W}_j, \mathbf{W}_j) dx + \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} (G_{jk} \frac{\mathbf{W}_k}{2}, \mathbf{W}_j) ds = 0, \quad (12)$$

et si  $C_j$  a une intersection non nulle avec la frontière  $\partial\Omega_h$  alors :

$$\int_{C_j} ((pG_0 + K) \mathbf{W}_j, \mathbf{W}_j) dx + \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} (G_{jk} \frac{\mathbf{W}_k}{2}, \mathbf{W}_j) ds + \frac{1}{2} \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} (|G_{jk}| \mathbf{W}_j, \mathbf{W}_j) ds = 0. \quad (13)$$

sachant que  $G_{jk} = G_{jk}^+ + G_{jk}^-$  et  $|G_{jk}| = G_{jk}^+ - G_{jk}^-$ .

En sommant les relations (12) et (13) sur toutes les cellules  $C_j$  on obtient :

$$\begin{aligned} & \frac{1}{2} \int_{\partial\Omega_h} (|G_{\mathbf{n}}| \mathbf{W}, \mathbf{W}) ds + \int_{\Omega_h} ((\nu G_0 + K) \mathbf{W}, \mathbf{W}) dx \quad + \\ & \frac{1}{2} \sum_{\Sigma_{jk} \in \mathcal{F}_h^{int}} \int_{\Sigma_{jk}} ((G_{jk} \mathbf{W}_k, \mathbf{W}_j) - (G_{jk} \mathbf{W}_j, \mathbf{W}_k)) ds = 0, \end{aligned} \quad (14)$$

puisque  $G_{kj} = -G_{jk}$  et où on a noté par  $\mathcal{F}_h^{int}$  l'ensemble des faces internes. En tenant compte du fait que (voir la proposition 12 de l'annexe 7.5) :

$$\begin{aligned} (G_{jk} \mathbf{W}_k, \mathbf{W}_j) - (G_{jk} \mathbf{W}_j, \mathbf{W}_k) &= (G_{jk} \mathbf{W}_k, \mathbf{W}_j) - \overline{(G_{jk} \mathbf{W}_k, \mathbf{W}_j)} \\ &= 2\Im((G_{jk} \mathbf{W}_k, \mathbf{W}_j)), \end{aligned}$$

et puisque les matrices  $|G_{\mathbf{n}}|$  et  $(\nu G_0 + K)$  sont hermitiennes (car symétriques réelles), on constate que la partie réelle de l'équation (14) est exactement la relation recherchée. Or, dans la mesure où les matrices  $|G_{\mathbf{n}}|$  et  $(\nu G_0 + K)$  sont respectivement positive et définie positive (car  $\nu > 0$ ), on conclut que la seule solution du problème homogène est la solution nulle.  $\blacksquare$

#### 2.4.2 Méthode de type Galerkin discontinu

On considère ici le cas  $m$  quelconque qui correspond à une approximation éléments finis polynomiale (cette méthode, plus particulièrement la variante  $m = 1$ , est étudiée plus en détail dans la section 2.4.2). Comme précédemment, l'approximation est donc discontinue, aucune relation de continuité étant imposée entre les éléments.

Si on introduit les espaces  $PH^s$  de fonctions  $H^s$  par morceaux, définis comme les fonctions qui sont localement  $H^s$  sur une partition de  $\Omega$  (par exemple sur les ouverts  $\Omega_j$  où les tenseurs  $\varepsilon$  et  $\mu$  sont constants), en imposant que  $\mathbf{W} \in (PH^s)^6$ , alors on garantit l'existence des traces sur l'ensembles des faces frontières ou internes.

**Lemme 2** *Pour chaque cellule  $C_j$  et pour chaque fonction test  $\mathbf{U} \in V_h$ , la solution  $\mathbf{W}_h \in V_h$  du problème homogène associé à (7) vérifie la relation suivante :*

$$\begin{aligned} \int_{C_j} ((pG_0 + K) \mathbf{W}_j, \mathbf{U}_j) dx &- \int_{C_j} (F(\mathbf{W}_j), \nabla \mathbf{U}_j) dx \\ &+ \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} (G_{jk} \mathbf{W}_k, \mathbf{U}_j) ds \\ &+ \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} (|G_{jk}| \mathbf{W}_j, \mathbf{U}_j) ds = 0. \end{aligned} \quad (15)$$

**Preuve** Si on intègre (7) sur une cellule  $C_j$  du maillage après avoir fait le produit scalaire avec une fonction test  $\mathbf{U}$  on obtient :

$$\int_{C_j} ((pG_0 + K)\mathbf{W}_j, \mathbf{U}_j) dx + \int_{C_j} (\operatorname{div} F(\mathbf{W}_j), \mathbf{U}_j) dx = 0.$$

Le théorème de Green-Riemann nous permet alors d'écrire :

$$\begin{aligned} \int_{C_j} ((pG_0 + K)\mathbf{W}_j, \mathbf{U}_j) dx & - \int_{C_j} (F(\mathbf{W}_j), \nabla \mathbf{U}_j) dx \\ & + \int_{\partial C_j} (F(\mathbf{W}_j) \cdot \mathbf{n}, \mathbf{U}_j) ds = 0. \end{aligned} \quad (16)$$

Comme dans le cas de la méthode volumes finis, on utilise une fonction de flux numérique centrée pour les interfaces internes et décentrée pour les interfaces frontières absorbantes et on note par  $G_{jk} = \sum_{l=1}^3 n_{jk}^{x_l} G_l$  où  $\mathbf{n}_{jk} = (n_{jk}^{x_1}, n_{jk}^{x_2}, n_{jk}^{x_3})^t$  est la normale unitaire à l'interface entre les cellules  $C_j$  et  $C_k$ , dirigée de  $C_j$  vers  $C_k$ . On obtient alors :

$$\begin{aligned} \int_{C_j} ((pG_0 + K)\mathbf{W}_j, \mathbf{U}_j) dx & - \int_{C_j} (F(\mathbf{W}_j), \nabla \mathbf{U}_j) dx \\ & + \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} (G_{jk} \left( \frac{\mathbf{W}_j + \mathbf{W}_k}{2} \right), \mathbf{U}_j) ds \\ & + \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} ((G_{jk}^+ \mathbf{W}_j + G_{jk}^- \mathbf{W}_k), \mathbf{U}_j) ds = 0. \end{aligned} \quad (17)$$

La condition absorbante exprimée faiblement correspond à  $G_{jk}^- \mathbf{W}_k = 0$ . Prenant cela en compte dans l'équation (17), on obtient :

$$\begin{aligned} \int_{C_j} ((pG_0 + K)\mathbf{W}_j, \mathbf{U}_j) dx & - \int_{C_j} (F(\mathbf{W}_j), \nabla \mathbf{U}_j) dx \\ & + \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} (G_{jk} \left( \frac{\mathbf{W}_j + \mathbf{W}_k}{2} \right), \mathbf{U}_j) ds \\ & + \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} (G_{jk}^+ \mathbf{W}_j, \mathbf{U}_j) ds = 0. \end{aligned} \quad (18)$$

Par ailleurs, le bilan de flux sur chaque cellule s'écrit :

$$\int_{\partial C_j} G_{jk} \mathbf{W}_j ds = \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} G_{jk} \mathbf{W}_j ds + \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} G_{jk} \mathbf{W}_j ds,$$

qui, utilisée dans (18), permet d'écrire :

$$\begin{aligned}
\int_{C_j} ((pG_0 + K)\mathbf{W}_j, \mathbf{U}_j) dx &= \int_{C_j} (F(\mathbf{W}_j), \nabla \mathbf{U}_j) dx \\
&+ \frac{1}{2} \int_{\partial C_j} (G_{jk} \mathbf{W}_j, \mathbf{U}_j) ds \\
&+ \frac{1}{2} \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} (G_{jk} \mathbf{W}_k, \mathbf{U}_j) ds \\
&+ \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} (G_{jk}^+ - \frac{G_{jk}}{2}) \mathbf{W}_j, \mathbf{U}_j) ds = 0,
\end{aligned} \tag{19}$$

ce qui équivaut à :

$$\begin{aligned}
2 \int_{C_j} ((pG_0 + K)\mathbf{W}_j, \mathbf{U}_j) dx &= 2 \int_{C_j} (F(\mathbf{W}_j), \nabla \mathbf{U}_j) dx \\
&+ \int_{\partial C_j} (G_{jk} \mathbf{W}_j, \mathbf{U}_j) ds \\
&+ \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} (G_{jk} \mathbf{W}_k, \mathbf{U}_j) ds \\
&+ \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} (|G_{jk}| \mathbf{W}_j, \mathbf{U}_j) ds = 0.
\end{aligned} \tag{20}$$

Ensuite, en réutilisant le théorème de Green-Riemann, mais cette fois-ci dans l'autre sens, on trouve :

$$\begin{aligned}
2 \int_{C_j} ((pG_0 + K)\mathbf{W}_j, \mathbf{U}_j) dx &= 2 \int_{C_j} (F(\mathbf{W}_j), \nabla \mathbf{U}_j) dx \\
&+ \int_{C_j} (\operatorname{div} F(\mathbf{W}_j), \mathbf{U}_j) + (F(\mathbf{W}_j), \nabla \mathbf{U}_j) dx \\
&+ \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} (G_{jk} \mathbf{W}_k, \mathbf{U}_j) ds \\
&+ \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} (|G_{jk}| \mathbf{W}_j, \mathbf{U}_j) ds = 0.
\end{aligned}$$

En réduisant les termes on obtient la relation désirée. ■

Si on somme les relations (2) sur toutes les cellules  $C_j$  on obtient la relation suivante :

$$a(\mathbf{W}, \mathbf{U}) + b(\mathbf{W}, \mathbf{U}) + c(\mathbf{W}, \mathbf{U}) = 0,$$

où  $a$ ,  $b$  et  $c$  sont les formes sesquilinéaires suivantes :

$$\begin{aligned}
a(\mathbf{W}, U) &= \int_{\Omega_h} ((pG_0 + K)\mathbf{W}, U) dx, \\
b(\mathbf{W}, U) &= \int_{\Omega_h} (-F(\mathbf{W}), \nabla U) dx, \\
c(\mathbf{W}, U) &= \sum_{\Sigma_{jk} \subset \mathcal{F}^{int}} \int_{\Sigma_{jk}} ((G_{jk}\mathbf{W}_k, U_j) - (G_{kj}\mathbf{W}_j, U_k)) ds \\
&\quad + \int_{\partial\Omega_h} (|G_n|\mathbf{W}, U) ds.
\end{aligned} \tag{21}$$

La proposition suivante va nous permettre de prouver que notre problème discret est bien posé.

**Proposition 1** *Les formes sesquilinéaires  $a$  et  $b$  définies dans (21) vérifient  $\Re(a(\mathbf{W}, U)) = \Re(b(\mathbf{W}, U))$ .*

**Preuve** Si on note  $\tilde{b}(\mathbf{W}, U) = b(\mathbf{W}, U) - a(\mathbf{W}, U)$  on peut voir facilement que :

$$\tilde{b}(\mathbf{W}, U) = \int_{\Omega_h} ((\operatorname{div} F(\mathbf{W}), U) - (F(\mathbf{W}), \nabla U)) dx,$$

est de partie réelle nulle en montrant que  $\tilde{b}(\mathbf{W}, U) = -\overline{\tilde{b}(U, \mathbf{W})}$ . ■

**Proposition 2** *Les formes sesquilinéaires  $a$ ,  $b$  et  $c$  définies dans (21) vérifient :*

$$\begin{aligned}
\Re(a(\mathbf{W}, \mathbf{W}) + b(\mathbf{W}, \mathbf{W}) + c(\mathbf{W}, \mathbf{W})) &= 2 \int_{\Omega_h} ((\nu G_0 + K)\mathbf{W}, \mathbf{W}) dx \\
&\quad + \int_{\partial\Omega_h} (|G_n|\mathbf{W}, \mathbf{W}) ds = 0,
\end{aligned} \tag{22}$$

et on peut conclure que l'unique solution du problème discret homogène est la solution nulle.

**Preuve** Dans un premier temps, sachant que  $p = i\omega + \nu$  et que  $(\nu G_0 + K)$  est une matrice symétrique réelle définie positive :

$$\Re(a(\mathbf{W}, \mathbf{W})) = \int_{\Omega_h} ((\nu G_0 + K)\mathbf{W}, \mathbf{W}) dx. \tag{23}$$

En outre :

$$(G_{jk}\mathbf{W}_k, \mathbf{W}_j) - (G_{jk}\mathbf{W}_j, \mathbf{W}_k) = (G_{jk}\mathbf{W}_k, \mathbf{W}_j) - \overline{(G_{jk}\mathbf{W}_k, \mathbf{W}_j)} = 2\Im((G_{jk}\mathbf{W}_k, \mathbf{W}_j)),$$

et on obtient que :

$$\Re(c(\mathbf{W}, \mathbf{W})) = \int_{\partial\Omega_h} (|G_n|\mathbf{W}, U) ds. \tag{24}$$

En sommant (23) et (24) et en tenant compte de la proposition 1, on retrouve l'égalité (10). Or, dans la mesure où  $|G_n|$  et  $(\nu G_0 + K)$  sont des matrices hermitiennes respectivement positive et définie positive (car  $\nu > 0$ ), ceci entraîne que  $\mathbf{W} = 0$  et la seule solution du problème homogène est la solution nulle. ■

### 3 Équations de Maxwell 1D

Pour dériver la version monodimensionnelle du système de Maxwell on prend comme direction de propagation de l'onde  $\mathbf{k} = (k, 0, 0)^t$  ce qui correspond à la direction  $x$ , et la polarisation du champ électrique telle que  $\mathbf{E} = (0, 0, E_z)^t$ . La polarisation du champ magnétique est déduite du produit vectoriel  $\mathbf{k} \times \mathbf{E}$  ce qui donne  $\mathbf{H} = (0, H_y, 0)^t$ . En supposant en plus que  $E_z$  et  $H_y$  sont des fonctions de  $x$  et  $t$  uniquement on obtient les équations de Maxwell en une dimension (on notera simplement  $E$  et  $H$  à la place de  $E_z$  et  $H_y$ ). On suppose aussi que  $\sigma = 0$  ce qui entraîne que  $J = 0$  :

$$\begin{cases} \frac{\partial}{\partial t}(\varepsilon E) - \frac{\partial H}{\partial x} = 0, \\ \frac{\partial}{\partial t}(\mu H) - \frac{\partial E}{\partial x} = 0. \end{cases} \quad (25)$$

On souhaite étudier le système (25) sur l'intervalle  $[a, b]$ . On doit donc définir les conditions aux limites en  $x = a$  et  $x = b$ . Rappelons qu'il est possible de résoudre les équations de Maxwell en champ total ou en champ diffracté. Pour formaliser simultanément les deux formulations, on introduit un paramètre  $\xi$  valant alternativement 0 si on souhaite résoudre (25) en champ diffracté (dans ce cas  $E$  désignera implicitement  $E^{\text{dif}}$ ) ou 1 si on souhaite résoudre ce système en champ total (et  $E$  désignera  $E^{\text{tot}}$ ).

Pour ce qui concerne les conditions aux limites, nous choisissons de considérer une condition métallique en  $x = b$  vérifiant  $E^{\text{tot}}(b) = 0$  et une condition absorbante de type Silver-Müller d'ordre 1 en  $x = a$  vérifiant  $[E^{\text{dif}} - zH^{\text{dif}}](a) = 0^2$ . Si on écrit ces conditions de manière à concilier les deux formulations, on a :

$$\begin{aligned} \begin{cases} E^{\text{tot}}(b) = 0 \\ [E^{\text{dif}} - zH^{\text{dif}}](a) = 0 \end{cases} &\Leftrightarrow \begin{cases} E(b) + (1 - \xi)E^{\text{inc}}(b) = 0 \\ [E - \xi E^{\text{inc}}](a) - z[H - \xi H^{\text{inc}}](a) = 0 \end{cases} \\ &\Leftrightarrow \begin{cases} E(b) = (\xi - 1)E^{\text{inc}}(b) \\ [E - zH](a) = \xi[E^{\text{inc}} - zH^{\text{inc}}](a). \end{cases} \end{aligned}$$

On procède maintenant à l'adimensionnement des équations (25) en effectuant les changements de variables suivants :

---

<sup>2</sup>Nous reviendrons plus loin sur l'origine de cette formulation



$$\begin{cases} \tilde{x} &= \frac{x}{L_0} \text{ et } \tilde{t} = \frac{c_0 t}{L_0} \\ \tilde{c} &= \frac{c}{c_0} = c_r \\ \tilde{E} &= \frac{E}{z_0 H_0} && \text{et idem pour } E^{\text{inc}}, \\ \tilde{H} &= \frac{H}{H_0} && \text{et idem pour } H^{\text{inc}}. \end{cases}$$

où  $c_0 = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}$ ,  $z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}}$ ,  $H_0$  est l'intensité du champ magnétique de référence et  $L_0$  la longueur de référence. Dans ce cas, nous avons :

$$\begin{cases} \partial_t(\varepsilon E) - \partial_x H = 0 \\ \partial_t(\mu H) - \partial_x E = 0 \\ E(b) = (\xi - 1)E^{\text{inc}}(b) \\ [E - zH](a) = \xi[E^{\text{inc}} - zH^{\text{inc}}](a) \end{cases} \Leftrightarrow \begin{cases} \partial_t(\varepsilon_0 \varepsilon_r z_0 H_0 \tilde{E}) - \partial_{\tilde{x}}(H_0 \tilde{H}) = 0 \\ \partial_t(\mu_0 \mu_r H_0 \tilde{H}) - \partial_{\tilde{x}}(z_0 H_0 \tilde{E}) = 0 \\ [z_0 H_0 \tilde{E}](b) = (\xi - 1)[z_0 H_0 \tilde{E}^{\text{inc}}](b) \\ [z_0 H_0 \tilde{E} - z_0 z_r H_0 \tilde{H}](a) = \xi[z_0 H_0 \tilde{E}^{\text{inc}} - z_0 z_r H_0 \tilde{H}^{\text{inc}}](a) \end{cases}$$

soit encore, en simplifiant par  $H_0$  et en tenant compte de l'adimensionnement de  $t$  et  $x$  :

$$\begin{cases} \partial_{\tilde{t}}(\frac{\varepsilon_0 \varepsilon_r z_0 c_0}{L_0} \tilde{E}) - \partial_{\tilde{x}}(\frac{\tilde{H}}{L_0}) = 0 \\ \partial_{\tilde{t}}(\frac{\mu_0 \mu_r c_0}{L_0} \tilde{H}) - \partial_{\tilde{x}}(\frac{z_0 \tilde{E}}{L_0}) = 0 \\ [z_0 H_0 \tilde{E}](b) = (\xi - 1)[z_0 H_0 \tilde{E}^{\text{inc}}](b) \\ [z_0 H_0 \tilde{E} - z_0 z_r H_0 \tilde{H}](a) = \xi[z_0 H_0 \tilde{E}^{\text{inc}} - z_0 z_r H_0 \tilde{H}^{\text{inc}}](a), \end{cases}$$

avec les notations classiques  $\partial_t = \frac{\partial}{\partial t}$  et  $\partial_x = \frac{\partial}{\partial x}$ . Pour conclure, après simplification, on obtient :

$$\begin{cases} \partial_{\tilde{t}}(\varepsilon_r) \tilde{E} - \partial_{\tilde{x}}(\tilde{H}) = 0 \\ \partial_{\tilde{t}}(\mu_r \tilde{H}) - \partial_{\tilde{x}}(\tilde{E}) = 0 \\ \tilde{E}(b) = (\xi - 1)\tilde{E}^{\text{inc}}(b) \\ [\tilde{E} - z_r \tilde{H}](a) = \xi[\tilde{E}^{\text{inc}} - z_r \tilde{H}^{\text{inc}}](a), \end{cases} \quad (26)$$

Dans la suite, pour alléger les notations, nous abandonnons le symbole  $\tilde{\phantom{x}}$  pour toutes les variables et nous notons simplement  $\varepsilon$  au lieu de  $\varepsilon_r$  (et idem pour  $\mu_r$ ). On obtient

alors un système formellement équivalent au système initial (25). On introduit les quantités suivantes :

$$W = \begin{pmatrix} E \\ H \end{pmatrix}, \quad \theta = \begin{pmatrix} \varepsilon & 0 \\ 0 & \mu \end{pmatrix}, \quad A = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad F(W) = AW. \quad (27)$$

Alors, les deux premières équations du système (26) peuvent se réécrire sous la forme conservative suivante :

$$\theta \frac{\partial W}{\partial t} + \frac{\partial F(W)}{\partial x} = 0. \quad (28)$$

Nous supposons maintenant que le champ  $W$  suit une dépendance harmonique en temps de pulsation  $\tilde{\omega} \in \mathbb{R}$  vérifiant  $\tilde{\omega} = \frac{\omega}{c_0}$ . Tout comme pour les autres variables, on assimile  $\tilde{\omega}$  à  $\omega$ , si bien que (28) devient :

$$i\omega\theta W(x) + \frac{\partial F(W)}{\partial x} = S, \quad (29)$$

où, comme on l'a montré dans la section 2.2, le terme source  $S$  est nul si on travaille en champ total et non nul dans le cas contraire.

### 3.1 Méthode de volumes finis

#### 3.1.1 Formulation du schéma

On discrétise l'intervalle  $[a, b]$  en  $N$  cellules  $C_j$  de taille  $\Delta x_j$  :  $C_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$  avec  $x_{\frac{1}{2}} = a$  et  $x_{N+\frac{1}{2}} = b$  (voir la figure 2).

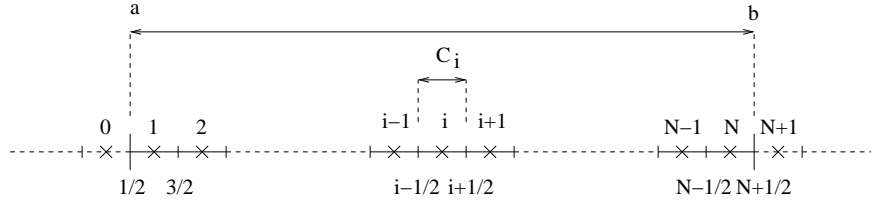


FIG. 2 – Discrétisation de l'intervalle  $[a, b]$

Les cellules  $C_0$  et  $C_{N+1}$  représentent des cellules fictives qui serviront dans ce qui suit à calculer les flux aux points  $x = a$  et  $x = b$ .

Pour une fonction vectorielle  $W$  on note  $W_j$  la valeur moyenne de  $W$  sur la cellule  $C_j$  :

$$W_j = \frac{1}{\Delta x_j} \int_{C_j} W(x) dx.$$

On notera également  $W_{j\pm\frac{1}{2}}$  la valeur approchée de  $W$  à l'interface entre les cellules  $C_j$  et  $C_{j\pm 1}$  ( $W_{j\pm\frac{1}{2}} \approx W(x_{j\pm\frac{1}{2}})$ ). En intégrant l'équation (29) sur la cellule  $\{C_j\}_{1\leq i\leq N}$ , on obtient :

$$\begin{aligned}
& \int_{C_j} (i\omega\theta W(x) + [F(W(x))]_x) dx = \int_{C_j} S(x) dx \\
\Leftrightarrow & \quad i\omega \int_{C_j} \theta W(x) dx + \int_{C_j} [F(W(x))]_x dx = \int_{C_j} S(x) dx \\
\Leftrightarrow & \quad i\omega \Delta x_j \theta_j W_j + \left( F(W(x_{j+\frac{1}{2}})) - F(W(x_{j-\frac{1}{2}})) \right) = \Delta x_j S_j \\
\Leftrightarrow & \quad i\omega \Delta x_j \theta_j W_j + \left( F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}} \right) = \Delta x_j S_j,
\end{aligned} \tag{30}$$

où  $\varepsilon_j$  et  $\mu_j$  sont les caractéristiques électromagnétiques associées à la cellule  $C_j$ , supposées constantes sur  $C_j$ . Par ailleurs,  $F_{j\pm\frac{1}{2}}$  est une approximation du flux  $F(W_{j\pm\frac{1}{2}})$  aux frontières des cellules. Pour les faces internes  $\{x_{j+\frac{1}{2}}\}_{1\leq j\leq N-1}$ , ces flux sont calculés de manière centrée [33] :

$$\begin{cases} F_{j-\frac{1}{2}} &= \frac{1}{2}(F_j + F_{j-1}) = A \left( \frac{W_j + W_{j-1}}{2} \right) \\ F_{j+\frac{1}{2}} &= \frac{1}{2}(F_j + F_{j+1}) = A \left( \frac{W_j + W_{j+1}}{2} \right) \end{cases}$$

Si on injecte ces expressions dans (30), on obtient pour  $j \in \{2, \dots, N-1\}$  :

$$\begin{aligned}
& i\omega \Delta x_j \theta_j W_j + F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}} = \Delta x_j S_j \\
\Leftrightarrow & \quad i\omega \Delta x_j \theta_j W_j + \frac{1}{2}(AW_{j+1} + AW_j) - \frac{1}{2}(AW_j + AW_{j-1}) = \Delta x_j S_j,
\end{aligned} \tag{31}$$

ce qui donne :

$$-AW_{j-1} + 2i\omega \Delta x_j \theta_j W_j + AW_{j+1} = 2\Delta x_j S_j. \tag{32}$$

Il s'agit maintenant de prendre en compte les conditions aux limites. Mais nous devons au préalable réaliser un calcul préliminaire qui nous sera utile pour le traitement de la condition absorbante de type Silver-Müller.

### 3.1.2 Diagonalisation de la matrice de flux

Le système de Maxwell temporel (26) est de nature hyperbolique donc diagonalisable. Cependant, cette diagonalisation doit s'opérer sur le système de Maxwell écrit sous forme conservative :

$$\frac{\partial Q}{\partial t} + \frac{\partial G(Q)}{\partial x} = 0, \tag{33}$$

avec :

$$Q = \begin{pmatrix} D \\ B \end{pmatrix} = \begin{pmatrix} \varepsilon & 0 \\ 0 & \mu \end{pmatrix} \begin{pmatrix} E \\ H \end{pmatrix} \quad \text{et} \quad G(Q) = MQ \quad \text{avec} \quad M = \begin{pmatrix} 0 & -\frac{1}{\mu} \\ -\frac{1}{\varepsilon} & 0 \end{pmatrix}$$

La diagonalisation de  $M$  s'écrit alors  $M = P\Lambda P^{-1}$  avec :

$$\Lambda = \begin{pmatrix} -c & 0 \\ 0 & c \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 1 \\ z & -z \end{pmatrix} \quad \text{et} \quad P^{-1} = \frac{1}{2} \begin{pmatrix} 1 & z^{-1} \\ 1 & -z^{-1} \end{pmatrix}.$$

D'autre part, si introduit les notations  $x^+ = \max(x, 0)$  et  $x^- = \min(x, 0)$  pour  $x \in \mathbb{R}$ , on a que  $x = x^+ + x^-$  et que  $|x| = x^+ - x^-$ . En appliquant ces notations aux éléments des matrices ci-dessus, on définit :

$$\Lambda^- = \begin{pmatrix} -c & 0 \\ 0 & 0 \end{pmatrix}, \quad \Lambda^+ = \begin{pmatrix} 0 & 0 \\ 0 & c \end{pmatrix} \quad \text{et} \quad |\Lambda| = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}.$$

De là, on définit aussi les matrices de flux positifs et négatifs, et on retrouve alors les relations exprimées plus haut. On a en effet :  $M^\pm = P\Lambda^\pm P^{-1}$  et  $|M| = P|\Lambda|P^{-1}$ . Ces matrices vérifient bien  $M = M^+ + M^-$  et  $|M| = M^+ - M^-$ . Elles valent respectivement :

$$M^+ = \frac{1}{2} \begin{pmatrix} c & -\frac{1}{\mu} \\ -\frac{1}{\varepsilon} & c \end{pmatrix}, \quad M^- = \frac{1}{2} \begin{pmatrix} -c & -\frac{1}{\mu} \\ -\frac{1}{\varepsilon} & -c \end{pmatrix} \quad \text{et} \quad |M| = |\Lambda| = cI_2.$$

### 3.1.3 Traitement des conditions aux limites

Le cadre est le suivant : on impose une condition absorbante au point  $x = a$ , une condition métallique au point  $x = b$ , et on souhaite envoyer une onde incidente  $W^{\text{inc}}$  qui se propage de  $x = a$  vers l'intérieur de  $[a, b]$ . Comme nous l'avons précisé précédemment, cette onde incidente devra être nécessairement solution des équations de Maxwell. Pour satisfaire cette condition, nous pouvons prendre une onde plane de la forme  $W^{\text{inc}}(x) = (\eta_e, \eta_h)e^{-kx}$  avec une condition sur le nombre d'onde  $k$ .

Pour traiter ces conditions aux limites, il convient de les respecter au sens faible. Ainsi, on approche les états aux points  $x = a$  et  $x = b$  par les formules suivantes :

- $W(a) \approx W_{\frac{1}{2}} = (W_0 + W_1)/2$
- $W(b) \approx W_{N+\frac{1}{2}} = (W_N + W_{N+1})/2$

Néanmoins, les cellules  $C_0$  et  $C_{N+1}$  sont fictives et on considère que les paramètres électromagnétiques  $\varepsilon$  et  $\mu$  correspondent à ceux de leurs cellules voisines. De plus, il faut choisir pour  $W_0$  et  $W_{N+1}$  des états fictifs respectivement associés aux cellules  $C_0$  et  $C_{N+1}$  de

sorte que  $W_{\frac{1}{2}}$  et  $W_{N+\frac{1}{2}}$  satisfassent respectivement les conditions absorbantes et métalliques. Ainsi, si on pose :

$$W_0 = \begin{pmatrix} 0 & z_1 \\ z_1^{-1} & 0 \end{pmatrix} W_1 + \xi \begin{pmatrix} 1 & -z_1 \\ -z_1^{-1} & 1 \end{pmatrix} W^{\text{inc}}, \quad (34)$$

$$W_{N+1} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} W_N + 2(\xi - 1) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} W^{\text{inc}}, \quad (35)$$

et :

$$A_1 = \begin{pmatrix} 1 & -z_1 \\ -z_1^{-1} & 1 \end{pmatrix} \quad \text{et} \quad A_N = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

alors :

$$\begin{aligned} A_1 \frac{W_0 + W_1}{2} &= \frac{1}{2} A_1 \left[ \begin{pmatrix} 1 & z_1 \\ z_1^{-1} & 1 \end{pmatrix} W_1 + \xi \begin{pmatrix} 1 & -z_1 \\ -z_1^{-1} & 1 \end{pmatrix} W^{\text{inc}} \right] \\ &= \frac{1}{2} A_1 \underbrace{\begin{pmatrix} 1 & z_1 \\ z_1^{-1} & 1 \end{pmatrix} W_1}_{=0} + \frac{\xi}{2} \underbrace{A_1^2}_{=2A_1} W^{\text{inc}} \\ &= \xi A_1 W^{\text{inc}}, \\ A_N \frac{W_N + W_{N+1}}{2} &= \frac{1}{2} A_N \left[ \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} W_N + 2(\xi - 1) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} W^{\text{inc}} \right] \\ &= \frac{1}{2} A_N \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} W_N}_{=0} + (\xi - 1) \underbrace{A_N^2}_{=A_N} W^{\text{inc}} \\ &= (\xi - 1) A_N W^{\text{inc}}. \end{aligned}$$

Compte tenu des propositions 4 et 5 énoncées dans l'annexe 7.1, les conditions limites sont bien respectées faiblement. Remarquons au passage que  $Z_1^{-1} M_1^+ \theta_1 = \frac{1}{2} A_1$  si bien que la condition au bord absorbant s'écrit aussi :

$$A_1 \frac{W_0 + W_1}{2} = \xi A_1 W^{\text{inc}} = 2\xi Z_1^{-1} M_1^+ \theta_1 W^{\text{inc}}$$

avec  $Z_1 = \begin{pmatrix} z_1^{-1} & 0 \\ 0 & z_1 \end{pmatrix}$ .

Nous pouvons passer à l'approximation des flux  $F_{\frac{1}{2}}$  et  $F_{N+\frac{1}{2}}$  associés respectivement aux points  $x = a$  et  $x = b$ . Tout comme pour les cellules internes, le calcul des flux aux frontières métallique et absorbante repose sur un schéma centré. Néanmoins, nous allons

voir que dans le cas de la frontière absorbante, la définition de  $W_{N+1}$  (34) va en fait conduire à l'utilisation d'un schéma décentré pour le calcul du flux en  $x = b$ .

Pour  $j = \frac{1}{2}$  le flux correspondant s'exprime par  $F_{\frac{1}{2}} = \frac{1}{2}(AW_0 + AW_1)$ . Compte tenu de la condition (34) et des paramètres électromagnétiques de  $C_0$  identifiés à ceux de  $C_1$ , on obtient :

$$\begin{aligned}
 F_{\frac{1}{2}} &= \frac{1}{2}A \left[ \begin{pmatrix} 0 & z_1 \\ z_1^{-1} & 0 \end{pmatrix} W_1 + \xi \begin{pmatrix} 1 & -z_1 \\ -z_1^{-1} & 1 \end{pmatrix} W^{\text{inc}} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} W_1 \right] \\
 &= \frac{1}{2}A \left[ \begin{pmatrix} 1 & z_1 \\ z_1^{-1} & 1 \end{pmatrix} W_1 + \xi \begin{pmatrix} 1 & -z_1 \\ -z_1^{-1} & 1 \end{pmatrix} W^{\text{inc}} \right] \\
 &= \frac{1}{2} \left[ \begin{pmatrix} -z_1^{-1} & -1 \\ -1 & -z_1 \end{pmatrix} W_1 + \xi \begin{pmatrix} z_1^{-1} & -1 \\ -1 & z_1 \end{pmatrix} W^{\text{inc}} \right] \\
 &= M_1^- \theta_1 W_1 + \xi M_1^+ \theta_1 W^{\text{inc}} = M_1^- Q_1 + \xi M_1^+ Q^{\text{inc}}
 \end{aligned}$$

On constate que le flux  $F_{\frac{1}{2}}$  est finalement évalué au moyen d'un schéma décentré similaire au schéma de Steger et Warming [37] souvent adopté pour le traitement des conditions aux limites à l'infini dans la résolution numérique des systèmes d'EDPs de la mécanique des fluides compressibles (voir par exemple [24]). Ce schéma permet de différencier les ondes entrantes des ondes sortantes à la frontière d'une cellule. Ainsi, la matrice des flux positifs (construite à l'aide des valeurs propres positives) permet de considérer les ondes allant de  $\{a\}$  vers  $\{b\}$ , c'est-à-dire allant de l'extérieur vers l'intérieur du domaine. Quant à la matrice des flux négatifs (construite à l'aide des valeurs propres négatives), elle permet de considérer les ondes allant de  $\{b\}$  vers  $\{a\}$ , c'est-à-dire allant de l'intérieur vers l'extérieur du domaine. La condition absorbante au point  $x = a$  exprime la non-réflexion des ondes provenant de la cellule  $C_1$  vers l'intérieur du domaine. Les seules ondes entrantes sont les ondes incidentes.

Le terme  $M_1^+ \theta_1 W^{\text{inc}}$  est donc une donnée du problème et, par souci de commodité, nous posons  $\gamma_1^{\text{inc}} = M_1^+ \theta_1 W^{\text{inc}}$ . Par suite :

$$F_{\frac{1}{2}} = M_1^- \theta_1 W_1 + \xi \gamma_1^{\text{inc}}. \quad (36)$$

En injectant (36) dans la dernière équation de (30), on obtient :

$$\begin{aligned}
 &i\omega \Delta x_1 \theta_1 W_1 + F_{\frac{3}{2}} - F_{\frac{1}{2}} = \Delta x_1 S_1 \\
 \Leftrightarrow &i\omega \Delta x_1 \theta_1 W_1 + \frac{1}{2}(AW_1 + AW_2) - (M_1^- \theta_1 W_1 + \xi \gamma_1^{\text{inc}}) = \Delta x_1 S_1 \\
 \Leftrightarrow &\left( i\omega \Delta x_1 I_2 + \frac{1}{2}A\theta_1^{-1} - M_1^- \right) \theta_1 W_1 + \frac{1}{2}AW_2 = \Delta x_1 S_1 + \xi \gamma_1^{\text{inc}} \\
 \Leftrightarrow &(2i\omega \Delta x_1 I_2 + M_1 - 2M_1^-) \theta_1 W_1 + AW_2 = 2(\Delta x_1 S_1 + \xi \gamma_1^{\text{inc}}).
 \end{aligned}$$

Or  $M_1 - 2M_1^- = (M_1^+ + M_1^-) - 2M_1^- = |M_1| = c_1 \mathbf{I}_2$  et on obtient donc pour la première cellule :

$$(2i\omega\Delta x_1 + c_1)\theta_1 W_1 + AW_2 = 2\left(\Delta x_1 S_1 + \xi\gamma_1^{\text{inc}}\right). \quad (37)$$

Similairement, le flux pour  $j = \frac{N}{2}$  s'écrit  $F_{N+\frac{1}{2}} = \frac{1}{2}(AW_N + AW_{N+1})$ . La cellule  $C_{N+1}$  étant fictive, nous allons imposer à cette cellule l'état miroir (35) pour la condition réfléchissante au point  $x = b$ . De plus, on identifie les paramètres électromagnétiques de  $C_{N+1}$  à ceux de  $C_N$ . On obtient alors :

$$\begin{aligned} F_{N+\frac{1}{2}} &= \frac{1}{2}(AW_N + AW_{N+1}) \\ &= \frac{1}{2}A \left[ W_N + \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} W_N + 2(\xi - 1) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} W^{\text{inc}} \right] \\ &= \frac{1}{2}A \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} W_N + (\xi - 1)A \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} W^{\text{inc}} \\ &= \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} W_N + (\xi - 1) \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} W^{\text{inc}}. \end{aligned}$$

Si on note  $P_N = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}$  et  $\gamma_N^{\text{inc}} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} W^{\text{inc}}$ , nous obtenons :

$$F_{N+\frac{1}{2}} = P_N W_N + (1 - \xi)\gamma_N^{\text{inc}}. \quad (38)$$

En injectant (38) dans (30), on en déduit :

$$\begin{aligned} i\omega\Delta x_N \theta_N W_N + F_{N+\frac{1}{2}} - F_{N-\frac{1}{2}} &= \Delta x_N S_N \\ \Leftrightarrow i\omega\Delta x_N \theta_N W_N + \left( P_N W_N + (1 - \xi)\gamma_N^{\text{inc}} \right) - \frac{1}{2}(AW_{N-1} + AW_N) &= \Delta x_N S_N \\ \Leftrightarrow -\frac{1}{2}AW_{N-1} + \left( i\omega\Delta x_N \theta_N + P_N - \frac{1}{2}A \right) W_N &= \Delta x_N S_N + (\xi - 1)\gamma_N^{\text{inc}} \\ \Leftrightarrow -AW_{N-1} + (2i\omega\Delta x_N \mathbf{I}_2 + (2P_N - A)\theta_N^{-1}) \theta_N W_N &= 2\left( \Delta x_N S_N + (\xi - 1)\gamma_N^{\text{inc}} \right). \end{aligned}$$

Or, si on note  $R_N = \begin{pmatrix} 0 & -\mu_N^{-1} \\ \varepsilon_N^{-1} & 0 \end{pmatrix}$ , alors  $(2P_N - A)\theta_N^{-1} = R_N$  et donc :

$$-AW_{N-1} + (2i\omega\Delta x_N \mathbf{I}_2 + R_N) \theta_N W_N = 2\left( \Delta x_N S_N + (\xi - 1)\gamma_N^{\text{inc}} \right). \quad (39)$$

Nous avons finalement évalué les flux aux interfaces de chaque cellule du maillage. En rassemblant les expressions (32), (37), et (39), nous nous ramenons à la formulation du système d'équations suivant :

$$\begin{cases} (2i\omega\Delta x_1 + c_1)\theta_1 W_1 + AW_2 = 2(\Delta x_1 S_1 + \xi\gamma_1^{\text{inc}}), \\ -AW_{j-1} + 2i\omega\Delta x_j \theta_j W_j + AW_{j+1} = 2\Delta x_j S_j, \text{ pour } 2 \leq j \leq N-1, \\ -AW_{N-1} + (2i\omega\Delta x_N I_2 + R_N)\theta_N W_N = 2(\Delta x_N S_N + (\xi-1)\gamma_N^{\text{inc}}). \end{cases} \quad (40)$$

### 3.1.4 Étude du problème discret

Dans cette section, on va d'abord écrire sous forme matricielle le système (40) en exhibant les propriétés des matrices qui interviennent. On va ensuite démontrer l'inversibilité du système linéaire résultant.

Avant d'exprimer le système (40) sous forme matricielle, nous allons le modifier de façon à obtenir des propriétés remarquables sur la matrice correspondante. On définit tout d'abord :

$$\Theta = \begin{pmatrix} \theta_1 & 0_2 & \dots & \dots & 0_2 \\ 0_2 & \theta_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \theta_{N-1} & 0_2 \\ 0_2 & \dots & \dots & 0_2 & \theta_N \end{pmatrix} \quad \text{avec} \quad \theta_j = \begin{pmatrix} \varepsilon_j & 0 \\ 0 & \mu_j \end{pmatrix}.$$

Comme  $\forall j \in \{1, \dots, N\}$ ,  $(\varepsilon_j, \mu_j) \in (\mathbb{R}_+^*)^2$ , il est clair que  $\Theta$  est une matrice définie positive. Si on pose  $D_j = 2\omega\Delta x_j \theta_j$ ,  $Z_j = c_j \theta_j = \begin{pmatrix} z_j^{-1} & 0 \\ 0 & z_j \end{pmatrix}$ , et si on note que  $\bar{R}_N \equiv R_N \theta_N = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ , alors on peut réécrire le système (40) sous la forme :

$$\begin{cases} (iD_1 + Z_1)W_1 + AW_2 = 2(\Delta x_1 S_1 + \xi\gamma_1^{\text{inc}}), \\ -AW_{j-1} + iD_j W_j + AW_{j+1} = 2\Delta x_j S_j, \text{ pour } 2 \leq j \leq N-1, \\ -AW_{N-1} + (iD_N + \bar{R}_N)W_N = 2(\Delta x_N S_N + (\xi-1)\gamma_N^{\text{inc}}), \end{cases} \quad (41)$$

et, sous forme matricielle :

$$\mathcal{A}W = b \quad (42)$$

où  $\mathcal{A}$ ,  $W$  et  $b$  sont donnés par :



$$\mathcal{A} = \begin{pmatrix} iD_1 + Z_1 & A & 0_2 & \dots & \dots & 0_2 \\ -A & iD_2 & A & \ddots & & \vdots \\ 0_2 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0_2 \\ \vdots & & \ddots & \ddots & iD_{N-1} & A \\ 0_2 & \dots & \dots & 0_2 & -A & iD_N + \bar{R}_N \end{pmatrix},$$

$$\mathbf{W} = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ \vdots \\ \vdots \\ W_{N-1} \\ W_N \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2(\Delta x_1 S_1 + \xi \gamma_1^{\text{inc}}) \\ 2\Delta x_2 S_2 \\ \vdots \\ \vdots \\ \vdots \\ 2\Delta x_{N-1} S_{N-1} \\ 2(\Delta x_N S_N + (\xi - 1)\gamma_N^{\text{inc}}) \end{pmatrix}.$$

Nous allons maintenant montrer que le système linéaire (42) est toujours inversible, ce qui revient à démontrer l'existence et l'unicité de la solution du problème discrétisé. Compte tenu des propriétés de la matrice  $\Theta$ , on est ramené à démontrer l'inversibilité de la matrice  $\mathcal{A}$ . Pour faciliter cette démonstration, on introduit les matrices  $\mathcal{D}$ ,  $\mathcal{Z}$  et  $\mathcal{R}$ , définies par :

$$\mathcal{D} = \begin{pmatrix} D_1 & 0_2 & \dots & 0_2 \\ 0_2 & D_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_2 \\ 0_2 & \dots & 0_2 & D_N \end{pmatrix}, \quad \mathcal{Z} = \begin{pmatrix} Z_1 & 0_2 & \dots & \dots & 0_2 \\ 0_2 & 0_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0_2 \\ 0_2 & \dots & \dots & 0_2 & 0_2 \end{pmatrix},$$

$$\mathcal{R} = \begin{pmatrix} 0_2 & A & 0_2 & \dots & 0_2 \\ -A & 0_2 & \ddots & \ddots & \vdots \\ 0_2 & \ddots & \ddots & \ddots & 0_2 \\ \vdots & \ddots & \ddots & 0_2 & A \\ 0_2 & \dots & 0_2 & -A & \bar{R}_N \end{pmatrix},$$

de telle sorte que  $\mathcal{A} = i\mathcal{D} + \mathcal{Z} + \mathcal{R}$ . On notera que  $\mathcal{Z}$  est une matrice réelle symétrique positive et que  $\mathcal{R}$  est une matrice antisymétrique. Nous pouvons alors démontrer le lemme suivant :

**Lemme 3** *La matrice  $\mathcal{A}$  du système linéaire (42) est inversible.*

**Preuve** Supposons que  $\mathcal{A}\mathbf{W} = 0$  pour  $\mathbf{W} \in \mathbb{C}^{2 \times N}$ . Montrons que  $W_j = 0, \forall j \in \llbracket 1, N \rrbracket$ . On constate dans un premier temps que :

$$\mathcal{A}\mathbf{W} = 0 \Rightarrow \mathbf{W}^* \mathcal{A}\mathbf{W} = 0 \Rightarrow \mathbf{W}^* (i\mathcal{D} + \mathcal{Z} + \mathcal{R})\mathbf{W} = 0 \Rightarrow \mathbf{W}^* (i\mathcal{D} + \mathcal{R})\mathbf{W} + \mathbf{W}^* \mathcal{Z}\mathbf{W} = 0.$$

où  $i\mathcal{D} + \mathcal{R}$  est une matrice antihermitienne. Par suite, d'après la proposition 12 de l'annexe 7.5,  $i\mathcal{D} + \mathcal{R} \in i\mathbb{R}$ . D'autre part, puisque  $\mathcal{Z}$  est une matrice réelle positive, on a que  $\mathbf{W}^* \mathcal{Z}\mathbf{W} \in \mathbb{R}_+$ . En résumé,  $\Re(\mathbf{W}^* \mathcal{A}\mathbf{W}) = \mathbf{W}^* \mathcal{Z}\mathbf{W}$  et  $i\Im(\mathbf{W}^* \mathcal{A}\mathbf{W}) = \mathbf{W}^* (i\mathcal{D} + \mathcal{R})\mathbf{W}$ . Donc  $\mathbf{W}^* \mathcal{A}\mathbf{W} = 0 \Rightarrow \mathbf{W}^* \mathcal{Z}\mathbf{W} = \mathbf{W}^* (i\mathcal{D} + \mathcal{R})\mathbf{W} = 0$ . Comme  $\mathbf{W}^* \mathcal{Z}\mathbf{W} = W_1^* Z_1 W_1$ , et comme  $Z_1$  est une matrice définie positive, alors nécessairement  $W_1 = 0$ . Par ailleurs :

$$\mathcal{A}\mathbf{W} = 0 \Rightarrow \forall j \in \{1, \dots, N\}, (\mathcal{A}\mathbf{W})_j = 0$$

En particulier, pour  $j = 1$ , on a :  $(iD_1 + Z_1)W_1 + AW_2 = 0$ . Or,  $W_1 = 0$ , donc  $AW_2 = 0$  et  $W_2 = 0$  car  $A$  est inversible. On procède ensuite par récurrence pour  $j \in \llbracket 3, N-1 \rrbracket$ . Supposons donc que  $W_{j-1} = W_j = 0$ . On a alors :

$$\mathcal{A}\mathbf{W} = 0 \Rightarrow (\mathcal{A}\mathbf{W})_j = 0 \Rightarrow -AW_{j-1} + iD_j W_j + AW_{j+1} = 0 \Rightarrow AW_{j+1} = 0 \Rightarrow W_{j+1} = 0.$$

En conclusion, l'hypothèse de récurrence est vraie et donc :

$$\mathcal{A}\mathbf{W} = 0 \Rightarrow \forall j \in \{1, N\} W_j = 0 \Rightarrow \mathbf{W} = 0$$

et  $\mathcal{A}$  est inversible. ■

**Remarque 1** Dans la démonstration précédente, on peut se rendre compte que c'est la matrice  $\mathcal{Z}$  dont les termes correspondent aux flux associés aux frontières absorbantes (le point  $x = a$  dans le cas présent) qui permet en partie de prouver l'inversibilité de la matrice  $\mathcal{A}$ . D'un point de vue purement algébrique, en l'absence de la matrice  $\mathcal{Z}$ , la partie réelle de la matrice  $\mathcal{A}$  serait antisymétrique et l'inversibilité de  $\mathcal{A}$  ne serait donc pas toujours garantie. D'un point de vue physique, les frontières absorbantes empêchent le champ électromagnétique d'être piégé et permettent donc d'éviter d'être confrontés à des modes de résonance selon que l'on choisisse la fréquence  $\omega$  en rapport avec la taille du domaine  $\Omega$ . Cela signifie que le problème discret est bien posé s'il comporte une condition absorbante, ce qui est mathématiquement vérifié dans le cadre continu [22].

## 3.2 Méthode de type Galerkin discontinu

### 3.2.1 Formulation du schéma

On se donne donc exactement le même cadre que celui adopté pour la méthode de volumes finis. Nous partons donc de l'équation :

$$i\omega\theta W(x) + \frac{\partial F(W)}{\partial x} = S. \quad (43)$$

Une formulation de type Galerkin discontinu (GD) repose sur l'introduction d'un ensemble de fonctions de base locales  $(\varphi_{jk})_{1 \leq k \leq d_j}$  pour chaque cellule  $(C_j)_{1 \leq j \leq N}$  où  $d_j$  re-

présente le nombre de degrés de liberté caractérisant l'approximation des variables d'état dans la cellule  $(C_j)$ . Dans chaque cellule, le champ  $W(x)$  est approché par une combinaison linéaire des fonctions de base locales (par exemple polynômiales de degré au plus  $m$ ), supposées linéairement indépendantes qui engendrent un espace noté  $\mathcal{P}_j = \mathbb{P}_m[C_j]$  :

$$W_j(x) \equiv W(x)|_{C_j} = \sum_{k=1}^{d_j} W_{jk} \varphi_{jk}(x), \quad (44)$$

où l'on a implicitement considéré ici le cas de fonctions de base scalaires. Les degrés de liberté locaux sont  $W_{jk} \in \mathbb{C}^2$ . Soit  $\varphi_j$  une fonction quelconque de  $\mathcal{P}_j$ . On multiplie (43) par  $\varphi_j$  et on intègre sur la cellule  $C_j$  :

$$\begin{aligned} & \int_{C_j} (i\omega \theta W(x) + [F(W(x))]_x) \varphi_j dx = \int_{C_j} S \varphi_j dx \\ \Leftrightarrow & i\omega \int_{C_j} \theta W(x) \varphi_j dx - \int_{C_j} F(W(x)) (\varphi_j)_x dx + \\ & F(W(x_{j+\frac{1}{2}})) \varphi_j(x_{j+\frac{1}{2}}) - F(W(x_{j-\frac{1}{2}})) \varphi_j(x_{j-\frac{1}{2}}) = \int_{C_j} S \varphi_j dx. \end{aligned} \quad (45)$$

Posons  $F_{j\pm\frac{1}{2}} = F(W(x_{j\pm\frac{1}{2}}))$  et  $\varphi_j^\pm = \varphi_j(x_{j\pm\frac{1}{2}})$ . Avec ces notations, l'équation (45) devient :

$$i\omega \int_{C_j} \theta W(x) \varphi_j dx - \int_{C_j} F(W(x)) (\varphi_j)_x dx + F_{j+\frac{1}{2}} \varphi_j^+ - F_{j-\frac{1}{2}} \varphi_j^- = \int_{C_j} S \varphi_j dx. \quad (46)$$

On pose  $\phi_j = (\varphi_{j1}, \varphi_{j2}, \dots, \varphi_{jd_j})$  la base locale de  $\mathcal{P}_j$ . Nous allons remplacer sur  $C_j$   $W$  par son approximation  $W_j \equiv W_j(x)$  défini par (44). Il s'agit maintenant d'injecter  $W_j$  dans l'équation (46). Notre problème est alors de chercher  $W_j \in \mathcal{P}_j$  tel que  $\forall \varphi_j \in \mathcal{P}_j$  :

$$i\omega \int_{C_j} \theta W_j \varphi_j dx - \int_{C_j} F(W_j(x)) (\varphi_j)_x dx + F_{j+\frac{1}{2}} \varphi_j^+ - F_{j-\frac{1}{2}} \varphi_j^- = \int_{C_j} S_j \varphi_j dx. \quad (47)$$

Comme pour la méthode VF, le calcul des flux  $F_{j\pm\frac{1}{2}}$  repose sur un schéma centré :

$$F_{j+\frac{1}{2}} = A \left( \frac{W_j^+ + W_{j+1}^-}{2} \right), \quad \text{avec } W_j^\pm = W_j(x_{j\pm\frac{1}{2}}), \quad (48)$$

ce qui conduit à :

$$\begin{aligned} \forall \varphi_i \in \mathcal{P}_j \quad : \quad i\omega \int_{C_j} \theta W_j \varphi_j dx & - \int_{C_j} AW_j(\varphi_j)_x dx \\ & + \frac{1}{2} (AW_j^+ + AW_{j+1}^-) \varphi_j^+ \\ & - \frac{1}{2} (AW_{j-1}^+ + AW_j^-) \varphi_j^- = \int_{C_j} S_j \varphi_j dx, \end{aligned} \quad (49)$$

c'est-à-dire :

$$\begin{aligned} \forall \varphi_j \in \mathcal{P}_j \quad : \quad i\omega \int_{C_j} \theta W_j \varphi_j dx & - \int_{C_j} AW_j(\varphi_j)_x dx \\ & + \frac{1}{2} AW_{j+1}^- \varphi_j^+ + \frac{1}{2} [AW_j \varphi_j]_{j-\frac{1}{2}}^{j+\frac{1}{2}} \\ & - \frac{1}{2} AW_{j-1}^+ \varphi_j^- = \int_{C_j} S_j \varphi_j dx. \end{aligned} \quad (50)$$

En utilisant la formule d'intégration par parties :

$$[AW_j \varphi_j]_{j-\frac{1}{2}}^{j+\frac{1}{2}} = \int_{C_j} AW_j(\varphi_j)_x dx + \int_{C_j} A(W_j)_x \varphi_j dx,$$

on obtient,  $\forall \varphi_j \in \mathcal{P}_j$  :

$$\begin{aligned} i\omega \theta_j \int_{C_j} W_j \varphi_j dx & + \frac{1}{2} \int_{C_j} (A(W_j)_x \varphi_j - AW_j(\varphi_j)_x) dx \\ & + \frac{1}{2} AW_{j+1}^- \varphi_j^+ - \frac{1}{2} AW_{j-1}^+ \varphi_j^- = \int_{C_j} S_j \varphi_j dx. \end{aligned} \quad (51)$$

où on a de plus supposé que les caractéristiques électromagnétiques  $\varepsilon_j$  et  $\mu_j$  sont constantes sur  $C_j$ .

### 3.2.2 Traitement des conditions aux limites

Tout comme dans le cas de la méthode de volumes finis, nous utilisons les états auxiliaires suivants :

$$W_0(x_{\frac{1}{2}}) = \begin{pmatrix} 0 & z_1 \\ z_1^{-1} & 0 \end{pmatrix} W_1(x_{\frac{1}{2}}) + \xi \begin{pmatrix} 1 & -z_1 \\ -z_1^{-1} & 1 \end{pmatrix} W^{\text{inc}}(x_{\frac{1}{2}}), \quad (52)$$

$$W_{N+1}(x_{N+\frac{1}{2}}) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} W_N(x_{N+\frac{1}{2}}) + 2(\xi - 1) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} W^{\text{inc}}(x_{N+\frac{1}{2}}), \quad (53)$$

qui sont telles que les conditions limites sont bien vérifiées de manière faible.

**Remarque 2** Il faut ici lever une ambiguïté concernant  $W^{inc}$  en définissant ce champ incident vis-à-vis des cellules frontières  $C_1$  et  $C_N$ . Ainsi, pour une formulation écrite par rapport à la cellule  $C_1$ , on notera  $(W^{inc})^-$  le champ  $W^{inc}$  pris au point  $\frac{1}{2}$ , et  $(W^{inc})^+$  le champ  $W^{inc}$  pris au point  $N + \frac{1}{2}$  pour une formulation écrite cette fois-ci par rapport à la cellule  $C_N$ .

**Condition absorbante au point  $x = a$ .** En utilisant (53) et en procédant comme pour la méthode de volumes finis :

$$\begin{aligned}
 AW_0^+ &= A \left( \begin{pmatrix} 0 & z_1 \\ z_1^{-1} & 0 \end{pmatrix} W_1^- + \xi \begin{pmatrix} 1 & -z_1 \\ -z_1^{-1} & 1 \end{pmatrix} (W^{inc})^- \right) \\
 &= \begin{pmatrix} -z_1^{-1} & 0 \\ 0 & -z_1 \end{pmatrix} W_1^- + \xi \begin{pmatrix} z_1^{-1} & -1 \\ -1 & z_1 \end{pmatrix} (W^{inc})^- \\
 &= -c_1 \theta_1 W_1^- + 2\xi M_1^+ \theta_1 (W^{inc})^- \\
 &= -c_1 Q_1^- + 2\xi M_1^+ (Q^{inc})^-.
 \end{aligned} \tag{54}$$

En réécrivant la formule (51) pour  $j = 1$  et en se servant de (54), on trouve :

$$\begin{aligned}
 &i\omega\theta_1 \int_{C_1} W_1 \varphi_1 dx + \frac{1}{2} \int_{C_1} (A(W_1)_x \varphi_1 - AW_1(\varphi_1)_x) dx \\
 &+ \frac{1}{2} AW_2^- \varphi_1^+ - \frac{1}{2} AW_0^+ \varphi_1^- = \int_{C_1} S_1 \varphi_1 dx \\
 \Leftrightarrow &i\omega\theta_1 \int_{C_1} W_1 \varphi_1 dx + \frac{1}{2} \int_{C_1} (A(W_1)_x \varphi_1 - AW_1(\varphi_1)_x) dx \\
 &+ \frac{1}{2} AW_2^- \varphi_1^+ - \frac{1}{2} (-c_1 Q_1^- + 2\xi M_1^+ (Q^{inc})^-) \varphi_1^- = \int_{C_1} S_1 \varphi_1 dx.
 \end{aligned}$$

On obtient donc :

$$\begin{aligned}
 \forall \varphi_1 \in \mathcal{P}_1, \quad \mathbb{B}\omega\theta_1 \int_{C_1} W_1 \varphi_1 dx &+ \frac{1}{2} \int_{C_1} (A(W_1)_x \varphi_1 - AW_1(\varphi_1)_x) dx \\
 &+ \frac{1}{2} c_1 Q_1^- \varphi_1^- + \frac{1}{2} AW_2^- \varphi_1^+ \\
 &= \int_{C_1} S_1 \varphi_1 dx + \xi M_1^+ (Q^{inc})^- \varphi_1^-.
 \end{aligned} \tag{55}$$

**Condition métallique au point  $x = b$ .** En utilisant (52) et en notant :

$$\rho_N = (\xi - 1) \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad \bar{R}_N = R_N \theta_N = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{et} \quad R_N = \begin{pmatrix} 0 & -\mu_N^{-1} \\ \varepsilon_N^{-1} & 0 \end{pmatrix}$$

alors :

$$\begin{aligned}
AW_{N+1}^- &= A \left( \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} W_N^+ + 2(\xi - 1) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} (W^{\text{inc}})^+ \right) \\
&= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} W_N^+ + 2(\xi - 1) \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} (W^{\text{inc}})^+ \\
&= \bar{R}_N W_N^+ - 2\rho_N (W^{\text{inc}})^+.
\end{aligned} \tag{56}$$

En réécrivant cette fois-ci la formule (51) pour  $j = N$  et en se servant de (56), on trouve :

$$\begin{aligned}
&i\omega\theta_N \int_{C_N} W_N \varphi_N dx + \frac{1}{2} \int_{C_N} (A(W_N)_x \varphi_N - AW_N(\varphi_N)_x) dx \\
&- \frac{1}{2} AW_{N-1}^+ \varphi_N^- + \frac{1}{2} AW_{N+1}^- \varphi_N^+ = \int_{C_N} S_N \varphi_N dx \\
\Leftrightarrow &i\omega\theta_N \int_{C_N} W_N \varphi_N dx + \frac{1}{2} \int_{C_N} (A(W_N)_x \varphi_N - AW_N(\varphi_N)_x) dx \\
&- \frac{1}{2} AW_{N-1}^+ \varphi_N^- + \frac{1}{2} (\bar{R}_N W_N^+ - 2\rho_N (W^{\text{inc}})^+) \varphi_N^+ = \int_{C_N} S_N \varphi_N dx.
\end{aligned}$$

On obtient donc pour tout  $\varphi_N \in \mathcal{P}_N$  :

$$\begin{aligned}
&i\omega\theta_N \int_{C_N} W_N \varphi_N dx + \frac{1}{2} \int_{C_N} (A(W_N)_x \varphi_N - AW_N(\varphi_N)_x) dx \\
&+ \frac{1}{2} \bar{R}_N W_N^+ \varphi_N^+ - \frac{1}{2} AW_{N-1}^+ \varphi_N^- = \int_{C_N} S_N \varphi_N dx + \rho_N (W^{\text{inc}})^+ \varphi_N^+.
\end{aligned} \tag{57}$$

### 3.2.3 Formulation matricielle

Posons  $\mathbb{W}_j = (W_{j1}, W_{j2}, \dots, W_{jd_j})$  une matrice complexe de taille  $2 \times d_j$ . La relation (44) s'écrit  $W_j = \mathbb{W}_j \phi_j^t$  où  $\phi_j = (\varphi_{j1}, \varphi_{j2}, \dots, \varphi_{jd_j})^t$ . En injectant cette expression dans (51) qui a été préalablement multipliée par 2, on obtient :

$$\begin{aligned}
\forall \varphi_j \in \mathcal{P}_j : \quad &2i\omega \int_{C_i} \theta_j \mathbb{W}_j \phi_j^t \varphi_j dx + \int_{C_j} (A\mathbb{W}_j(\phi_j)_x^t \varphi_j - A\mathbb{W}_j \phi_j^t (\varphi_j)_x) dx \\
&+ A\mathbb{W}_{j+1}(\phi_{j+1}^-)^t \varphi_j^+ - A\mathbb{W}_{j-1}(\phi_{j-1}^+)^t \varphi_j^- = 2 \int_{C_j} \mathbb{S}_j \phi_j^t \varphi_j dx,
\end{aligned} \tag{58}$$

Introduisons maintenant les notations suivantes :

$$\begin{cases} \Phi_j &= \int_{C_j} \phi_j^t \phi_j dx, \\ \Phi'_j &= \int_{C_j} (\phi_j^t (\phi_j)_x - (\phi_j)_x^t \phi_j) dx, \\ \Phi_j^\pm &= (\phi_j^\pm)^t \phi_j^\pm = \phi_j^t(x_{j\pm\frac{1}{2}}) \phi_j^t(x_{j\pm\frac{1}{2}}), \\ \Phi_{j\pm 1,j} &= (\phi_{j\pm 1}^\mp)^t \phi_j^\pm = \phi_{j\pm 1}^t(x_{j\pm\frac{1}{2}}) \phi_j(x_{j\pm\frac{1}{2}}). \end{cases} \quad (59)$$

On peut vérifier que les relations ci-dessous s'appliquent :

$$\begin{cases} \Phi_j^t &= \Phi_j, \\ (\Phi'_j)^t &= -\Phi'_j, \\ (\Phi_j^\pm)^t &= \Phi_j^\pm, \\ (\Phi_{j,j\pm 1})^t &= \Phi_{j\pm 1,j}. \end{cases} \quad (60)$$

Clairement la matrice de masse locale  $\Phi_j$  est symétrique définie positive,  $\Phi_j^\pm$  est symétrique, alors que  $\Phi'_j$  est antisymétrique. Maintenant, en prenant pour  $\varphi_j$  successivement toutes les fonctions de base  $\varphi_{jk}$  pour  $1 \leq k \leq d_j$ , on peut reformuler les équations (55), (51) et (57) comme suit :

$$2i\omega\theta_1\mathbb{W}_1\Phi_1 + A\mathbb{W}_1(\Phi'_1)^t + A\mathbb{W}_2\Phi_{2,1} + Z_1\mathbb{W}_1\Phi_1^- = 2\left(\mathbb{S}_1\Phi_1 + \xi M_1^+\theta_1\mathbb{W}^{\text{inc}}\Phi_1^-\right), \quad (61)$$

$$2 \leq j \leq N-1, \quad 2i\omega\theta_j\mathbb{W}_j\Phi_j + A\mathbb{W}_j(\Phi'_j)^t + A\mathbb{W}_{j+1}\Phi_{j+1,j} - A\mathbb{W}_{j-1}\Phi_{j-1,j} = 2\mathbb{S}_j\Phi_j, \quad (62)$$

$$2i\omega\theta_N\mathbb{W}_N\Phi_N + A\mathbb{W}_N(\Phi'_N)^t - A\mathbb{W}_{N-1}\Phi_{N-1,N} + \bar{R}_N\mathbb{W}_N\Phi_N^+ = 2\left(\mathbb{S}_N\Phi_N + \rho_N\mathbb{W}^{\text{inc}}\Phi_N^+\right), \quad (63)$$

où  $Z_1 = c_1\theta_1 = \begin{pmatrix} z_1^{-1} & 0 \\ 0 & z_1 \end{pmatrix}$ .

Pour aboutir à un système linéaire à partir des équations (61), (62), et (63) nous voyons qu'il faut convertir notre inconnue matricielle  $\mathbb{W} \in \mathbf{M}_{2,d_j}(\mathbb{C})$  pour tout  $j$  dans  $\llbracket 1; N \rrbracket$  en un vecteur de  $\mathbb{C}^{d_j \times 2}$  ou de  $\mathbb{C}^{2 \times d_j}$ . Nous utiliserons pour ce faire, un procédé de vectorisation, noté  $\aleph$ , qui transforme la matrice  $\mathbb{W}_j = (W_{j1}, W_{j2}, \dots, W_{jd_j})$  en un vecteur colonne  $\mathbf{W}_j = (W_{j1}, W_{j2}, \dots, W_{jd_j})^t \in \mathbb{C}^{2 \times d_j}$ . L'application  $\aleph$  répond à la proposition 14 de l'annexe 7.7.1. Elle fait intervenir un produit de Kronecker  $\otimes$ , lui-même explicité dans cette proposition, et vérifie alors  $\aleph(\mathbb{W}_j) = \mathbf{W}_j$ . D'autre part, sachant que l'on considère une formulation où  $m$  est fixé pour toutes les cellules, nous pouvons écrire que  $d_j = d, \forall j \in \llbracket 1; N \rrbracket$ .

En appliquant maintenant  $\aleph$  à l'équation (62), nous obtenons par linéarité que (62) est successivement équivalente à :

$$2i\omega \aleph(\theta_j \mathbb{W}_j \Phi_j) + \aleph(A \mathbb{W}_j (\Phi'_j)^t) + \aleph(A \mathbb{W}_{j+1} \Phi_{j+1,j}^+) - \aleph(A \mathbb{W}_{j-1} \Phi_{j-1,j}^-) = 2\aleph(S_j \Phi_j) \Leftrightarrow$$

$$(2i\omega \Phi_j \otimes \theta_j + \Phi'_j \otimes A) \mathbf{W}_j + (\Phi_{j,j+1}^+ \otimes A) \mathbf{W}_{j+1} - (\Phi_{j,j-1}^- \otimes A) \mathbf{W}_{j-1} = 2(\Phi_j \otimes \mathbf{I}_2) \mathbf{S}_j. \quad (64)$$

En procédant de même pour les équations (61) (63), on trouve que (61) est équivalente à :

$$(2i\omega \Phi_1 \otimes \theta_1 + \Phi'_1 \otimes A + \Phi_1^- \otimes Z_1) \mathbf{W}_1 + (\Phi_{1,2} \otimes A) \mathbf{W}_2 = 2(\Phi_1 \otimes \mathbf{I}_2) \mathbf{S}_j$$

$$+ 2\xi(\Phi_1^- \otimes M_1^+) \theta_1 W^{\text{inc}}, \quad (65)$$

et (63) est équivalente à :

$$(2i\omega(\Phi_N \otimes \theta_N) + (\Phi'_N \otimes A) + (\Phi_N^+ \otimes \bar{R}_N)) \mathbf{W}_N$$

$$- (\Phi_{N,N-1} \otimes A) \mathbf{W}_{N-1} = 2(\Phi_N \otimes \mathbf{I}_2) \mathbf{S}_N + 2(\Phi_N^+ \otimes \rho_N) W^{\text{inc}} \quad (66)$$

A première vue, il n'est pas évident que (64), (65) et (66) permettent d'aboutir à une matrice qui conserve la structure de la matrice issue du schéma volumes finis. Cela dit, grâce à certaines propriétés conférées par le produit de Kronecker, nous verrons que les deux matrices possèdent en effet la même structure. Ces propriétés sont énoncées dans l'annexe 7.7.2. Nous définissons ainsi les matrices suivantes :

$$\mathcal{D} = \begin{pmatrix} 2\omega \Phi_1 \otimes \theta_1 & 0_{2d} & \dots & 0_{2d} \\ 0_{2d} & 2\omega \Phi_2 \otimes \theta_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{2d} \\ 0_{2d} & \dots & 0_{2d} & 2\omega \Phi_N \otimes \theta_N \end{pmatrix}, \quad \mathcal{Z} = \begin{pmatrix} \Phi_1^- \otimes Z_1 & 0_{2d} & \dots & 0_{2d} \\ 0_{2d} & 0_{2d} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{2d} \\ 0_{2d} & \dots & 0_{2d} & 0_{2d} \end{pmatrix},$$

$$\mathcal{R} = \begin{pmatrix} \Phi'_1 \otimes A & \Phi_{1,2} \otimes A & 0_{2d} & \dots & 0_{2d} \\ -\Phi_{2,1} \otimes A & \Phi'_2 \otimes A & \Phi_{2,3} \otimes A & \ddots & \vdots \\ 0_{2d} & \ddots & \ddots & \ddots & 0_{2d} \\ \vdots & \ddots & -\Phi_{N-1,N-2} \otimes A & \Phi'_{N-1} \otimes A & \Phi_{N-1,N} \otimes A \\ 0_{2d} & \dots & 0_{2d} & -\Phi_{N,N-1} \otimes A & \Phi'_N \otimes A + \Phi_N^+ \otimes \bar{R}_N \end{pmatrix}.$$

La matrice  $\mathcal{A}$  résultant du schéma (64)-(65)-(66) peut encore s'écrire encore sous la forme  $\mathcal{A} = i\mathcal{D} + \mathcal{Z} + \mathcal{R}$ , comme dans le cas de la méthode de volumes finis.

Dans ce qui suit, on explicite ces matrices dans le cas particulier où l'on prend des fonctions de base  $\mathbb{P}_1$  puis on démontre l'inversibilité du système linéaire résultant.



### 3.2.4 Formulation $P_1$

On se place maintenant dans le cas concret des fonctions affines par morceaux, i.e.  $\mathcal{P}_i = \mathbb{P}_1[C_j]$ . Dans ce cas  $\phi_j = (\varphi_{j1}, \varphi_{j2})$ . Pour le calcul numérique des différents termes de (59), on se place dans le contexte habituel des éléments finis, c'est-à-dire que l'on raisonne sur une cellule de référence  $\hat{C}$  qui est ici l'intervalle  $[0,1]$ . On prend sur cet élément les fonctions de base  $\hat{\varphi}_1(\xi) = 1 - \xi$  et  $\hat{\varphi}_2(\xi) = \xi$ . Les fonctions de base sur  $C_j$  sont alors  $\varphi_{j1}(x) = \hat{\varphi}_1 \left[ (x - x_{j-\frac{1}{2}})/\Delta x_j \right]$  et  $\varphi_{j2}(x) = \hat{\varphi}_2 \left[ (x - x_{j-\frac{1}{2}})/\Delta x_j \right]$ . On a, pour  $1 \leq j \leq N$  :

$$\left\{ \begin{array}{l} \Phi_j = \int_{C_j} \phi_j^t \phi_j dx = \Delta x_j \int_{\hat{C}} \hat{\phi}^t \hat{\phi} d\hat{x} = \frac{\Delta x_j}{6} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \\ \Phi'_j = \int_{C_j} (\phi_j^t (\phi_j)_x - (\phi_j)_x^t \phi_j) dx = \Delta x_j \int_{\hat{C}} \frac{1}{\Delta x_j} \left( \hat{\phi}^t (\hat{\phi})_x - (\hat{\phi})_x^t \hat{\phi} \right) d\hat{x} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \\ \Phi_j^+ = (\phi_j^+)^t \phi_j^+ = \phi_j^t(x_{j+\frac{1}{2}}) \phi_j(x_{j+\frac{1}{2}}) = \hat{\phi}^t(1) \hat{\phi}(1) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \\ \Phi_j^- = (\phi_j^-)^t \phi_j^- = \phi_j^t(x_{j-\frac{1}{2}}) \phi_j(x_{j-\frac{1}{2}}) = \hat{\phi}^t(0) \hat{\phi}(0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \end{array} \right.$$

D'autre part, on trouve aussi que :

$$\begin{aligned} \forall j, 1 \leq j \leq N-1, \quad \Phi_{j,j+1} &= (\phi_j^+)^t \phi_{j+1}^- = \phi_j^t(x_{j+\frac{1}{2}}) \phi_{j+1}(x_{j+\frac{1}{2}}) = \hat{\phi}^t(1) \hat{\phi}(0) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \\ \forall j, 2 \leq j \leq N, \quad \Phi_{j,j-1} &= (\phi_j^-)^t \phi_{j-1}^+ = \phi_j^t(x_{j-\frac{1}{2}}) \phi_{j-1}(x_{j-\frac{1}{2}}) = \hat{\phi}^t(0) \hat{\phi}(1) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Enfin, si on multiplie (64) par 3 pour éviter les termes fractionnaires, et si on formule un système linéaire  $\mathcal{A}W = b$  avec  $(\mathcal{A}, b) \in \mathbf{M}_{4 \times N}(\mathbb{C}) \times \mathbb{C}^{4 \times N}$ , alors la matrice  $\mathcal{A}$  est définie par blocs par :

$$\begin{aligned}
\bullet \mathcal{A}_{1,1} &= \left( \begin{array}{cc|cc} 2i\omega\Delta x_1\varepsilon_1 + 3z_1^{-1} & 0 & i\omega\Delta x_1\varepsilon_1 & -3 \\ 0 & 2i\omega\Delta x_1\mu_1 + 3z_1 & -3 & i\omega\Delta x_1\mu_1 \\ \hline i\omega\Delta x_1\varepsilon_1 & 3 & 2i\omega\Delta x_1\varepsilon_1 & 0 \\ 3 & i\omega\Delta x_1\mu_1 & 0 & 2i\omega\Delta x_1\mu_1 \end{array} \right), \\
\bullet \mathcal{A}_{1,2} &= \left( \begin{array}{cc|c} 0_2 & & 0_2 \\ \hline 0 & -3 & 0_2 \\ -3 & 0 & 0_2 \end{array} \right), \\
\bullet \mathcal{A}_{j,j-1} &= \left( \begin{array}{cc|cc} & & 0 & 3 \\ 0_2 & & 3 & 0 \\ \hline 0_2 & & 0_2 & \end{array} \right), \quad \forall j \in \llbracket 2; N-1 \rrbracket, \\
\bullet \mathcal{A}_{jj} &= \left( \begin{array}{cc|cc} 2i\omega\Delta x_j\varepsilon_j & 0 & i\omega\Delta x_j\varepsilon_j & -3 \\ 0 & 2i\omega\Delta x_j\mu_j & -3 & i\omega\Delta x_j\mu_j \\ \hline i\omega\Delta x_j\varepsilon_j & 3 & 2i\omega\Delta x_j\varepsilon_j & 0 \\ 3 & i\omega\Delta x_j\mu_j & 0 & 2i\omega\Delta x_j\mu_j \end{array} \right), \quad \forall j \in \llbracket 2; N-1 \rrbracket, \\
\bullet \mathcal{A}_{j,j+1} &= \left( \begin{array}{cc|c} 0_2 & & 0_2 \\ \hline 0 & -3 & 0_2 \\ -3 & 0 & 0_2 \end{array} \right), \quad \forall j \in \llbracket 2; N-1 \rrbracket, \\
\bullet \mathcal{A}_{N,N-1} &= \left( \begin{array}{cc|cc} & & 0 & 3 \\ 0_2 & & 3 & 0 \\ \hline 0_2 & & 0_2 & \end{array} \right), \\
\bullet \mathcal{A}_{N,N} &= \left( \begin{array}{cc|cc} 2i\omega\Delta x_N\varepsilon_N & 0 & i\omega\Delta x_N\varepsilon_N & -3 \\ 0 & 2i\omega\Delta x_N\mu_N & -3 & i\omega\Delta x_N\mu_N \\ \hline i\omega\Delta x_N\varepsilon_N & 3 & 2i\omega\Delta x_N\varepsilon_N & -3 \\ 3 & i\omega\Delta x_N\mu_N & 3 & 2i\omega\Delta x_N\mu_N \end{array} \right).
\end{aligned}$$

Ces matrices ont les mêmes caractéristiques que celles du schéma VF, à savoir :

- $\mathcal{D}$  est diagonale par blocs et symétrique définie positive,
- $\mathcal{Z}$  est diagonale positive,
- $\mathcal{R}$  est antisymétrique.

Il est montré dans la proposition 15 de l'annexe 7.7.4 qu'une matrice résultant de l'assemblage de fonctions  $\mathbb{P}_m$  possède la même structure. Pour autant, l'inversibilité ne sera pas forcément assurée. Néanmoins, nous allons voir dans ce qui suit que pour notre formulation d'ordre 1, l'inversibilité de la matrice est vérifiée.

### 3.2.5 Étude du problème discret

Nous constatons malheureusement que  $\forall j \in \llbracket 1; N-1 \rrbracket$  la matrice  $\mathcal{A}_{j(j+1)}$  n'est pas inversible. Ainsi, il n'est pas possible de reproduire exactement la démonstration de l'inversibilité de la matrice associée à la méthode de volumes finis. En revanche, nous allons faire une démonstration similaire dont l'idée sera de perpétuer une relation de récurrence non plus par cellule mais par degré de liberté. Nous considérons que  $\mathbf{W} = (W_1^t, W_2^t, \dots, W_N^t)^t$  où pour  $j \in \llbracket 1; N \rrbracket$   $W_j^t = (W_j^1, W_j^2)^t$  avec  $W_j^{1,2} \in \mathbb{C}^2$ .

**Lemme 4** *La matrice  $\mathcal{A}$  définie par  $\mathcal{A} = i\mathcal{D} + \mathcal{Z} + \mathcal{R}$ , caractérisant le schéma Galerkin Discontinu 1D en formulation  $\mathbb{P}_1$ , est inversible.*

**Preuve** Montrons par récurrence sur  $j$  pour  $j \in \llbracket 1, N-1 \rrbracket$  que si  $\mathcal{A}\mathbf{W} = 0$  pour  $\mathbf{W} \in \mathbb{C}^{4 \times N}$  alors  $W_j = W_{j+1} = 0$ .

1. Pour le premier rang de la récurrence, nous avons :

$$\mathcal{A}\mathbf{W} = 0 \Rightarrow \mathbf{W}^* \mathcal{A}\mathbf{W} = 0 \Rightarrow \mathbf{W}^* (i\mathcal{D} + \mathcal{Z} + \mathcal{R})\mathbf{W} = 0 \Rightarrow \mathbf{W}^* (i\mathcal{D} + \mathcal{R})\mathbf{W} + \mathbf{W}^* \mathcal{Z}\mathbf{W} = 0.$$

La matrice  $\mathcal{Z}$  est hermitienne donc  $\mathbf{W}^* \mathcal{Z}\mathbf{W} \in \mathbb{R}$ . D'autre part, reprenant les arguments de la démonstration réalisée pour la méthode de volumes finis,  $(i\mathcal{D} + \mathcal{R})$  est antihermitienne et donc  $\mathbf{W}^* (i\mathcal{D} + \mathcal{R})\mathbf{W} \in i\mathbb{R}$ . Ceci entraîne donc que  $\Re(\mathbf{W}^* \mathcal{A}\mathbf{W}) = \mathbf{W}^* \mathcal{Z}\mathbf{W}$  et  $\Im(\mathbf{W}^* \mathcal{A}\mathbf{W}) = \mathbf{W}^* (i\mathcal{D} + \mathcal{R})\mathbf{W}$ .

En conséquence,  $\mathbf{W}^* \mathcal{A}\mathbf{W} = 0 \Rightarrow \mathbf{W}^* \mathcal{Z}\mathbf{W} = \mathbf{W}^* (i\mathcal{D} + \mathcal{R})\mathbf{W} = 0$ . Si on s'intéresse plus particulièrement à  $\mathbf{W}^* \mathcal{Z}\mathbf{W}$ , nous obtenons en développant que  $\mathbf{W}^* \mathcal{Z}\mathbf{W} = W_1^* \mathcal{Z}_{11} W_1 = W_1^* (\Phi_1^- \otimes Z_1) W_1 = 0$ . De là on tire  $W_1^1 = 0$ .

Maintenant, nous avons aussi :

$$\begin{aligned} \mathcal{A}\mathbf{W} = 0 &\Rightarrow \forall j \in \llbracket 1; N \rrbracket, (\mathcal{A}\mathbf{W})_j = 0 \quad (\text{et en particulier pour } j = 1) \\ &\Rightarrow A_{11}W_1 + A_{12}W_2 = 0 \\ &\Rightarrow \mathcal{A}_{11}(0, (W_1^2)^t)^t + A_{12}W_2 = 0 \\ &\Rightarrow \begin{cases} \begin{pmatrix} i\omega\Delta x_1 \varepsilon_1 & -3 \\ -3 & i\omega\Delta x_1 \mu_1 \end{pmatrix} W_1^2 = 0 \\ \begin{pmatrix} i\omega\Delta x_1 \varepsilon_1 & 0 \\ 0 & i\omega\Delta x_1 \mu_1 \end{pmatrix} W_1^2 + \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix} W_2^1 = 0. \end{cases} \end{aligned}$$

Notons  $C = \begin{pmatrix} i\omega\Delta x_1 \varepsilon_1 & -3 \\ -3 & i\omega\Delta x_1 \mu_1 \end{pmatrix}$ . Puisque  $\det C = -(\frac{\omega^2 \Delta x_1^2}{c_1^2} + 9) < 0$ , alors  $C$  est inversible si bien que  $W_1^2 = 0$ . Combinant le fait que  $W_1^2 = 0$  et l'autre égalité, alors il est évident que  $W_2^1 = 0$ .

Nous avons donc montré que  $\mathcal{A}\mathbf{W} = 0 \Rightarrow W_1 = 0$  et  $W_2^1 = 0$ .

2. Supposons maintenant que la relation est vérifiée au rang  $j$ , c'est-à-dire que  $W_{j-1} = 0$  pour  $j \in \llbracket 2, N-1 \rrbracket$  et que  $W_j^1 = 0$ . Nous devons montrer que  $W_j = 0$  et que  $W_{j+1}^1 = 0$ . Mais vu que  $W_j^1 = 0$ , alors il faut juste prouver que  $W_j^2 = 0$  et  $W_{j+1}^1 = 0$  sachant que :

$$\begin{aligned}
\mathcal{A}\mathbf{W} = 0 &\Rightarrow \forall j \in \llbracket 1, N \rrbracket, (\mathcal{A}\mathbf{W})_j = 0 \quad (\text{et en particulier pour } j \in \llbracket 2, N-1 \rrbracket) \\
&\Rightarrow \mathcal{A}_{j(j-1)} W_{j-1} + \mathcal{A}_{jj} W_j + \mathcal{A}_{j(j+1)} W_{j+1} = 0 \\
&\Rightarrow \mathcal{A}_{jj}(0, (W_j^2)^t)^t + \mathcal{A}_{j(j+1)} W_{j+1} = 0 \\
&\Rightarrow \begin{cases} \begin{pmatrix} 2i\omega\Delta x_j \varepsilon_j & -3 \\ -3 & 2i\omega\Delta x_j \mu_j \end{pmatrix} W_j^2 = 0 \\ \begin{pmatrix} i\omega\Delta x_j \varepsilon_j & 0 \\ 0 & i\omega\Delta x_j \mu_j \end{pmatrix} W_j^2 + \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix} W_{j+1}^1 = 0. \end{cases}
\end{aligned}$$

Avec ce système d'égalités, on reprend à la lettre le cheminement du 1. pour trouver au final que  $W_j = 0$  et que  $W_{j+1}^1 = 0$ .

3. Il ne reste plus qu'à montrer que  $W_N^2 = 0$  sachant que pour  $j \in \llbracket 1, N-1 \rrbracket$   $W_j = 0$  et que  $W_N^1 = 0$ . En développant la dernière ligne du système et en tenant compte de ces affirmations, on trouve :

$$\mathcal{A}\mathbf{W} = 0 \Rightarrow \begin{cases} \begin{pmatrix} i\omega\Delta x_N \varepsilon_N & -3 \\ -3 & i\omega\Delta x_N \mu_N \end{pmatrix} W_N^2 = 0, \\ \begin{pmatrix} 2i\omega\Delta x_N \varepsilon_N & -3 \\ 3 & 2i\omega\Delta x_N \mu_N \end{pmatrix} W_N^2 = 0. \end{cases} \quad (67)$$

On aboutit comme précédemment à  $W_N^2 = 0$ . ■

## 4 Équations de Maxwell 3D

Pour des soucis d'implémentation, nous souhaitons traiter de front la résolution du système en champ diffracté et en champ total en introduisant le paramètre  $\xi$  qui permet de distinguer un calcul en champ diffracté ( $\xi = 0$ ) d'un calcul en champ total ( $\xi = 1$ ). Rappelons que :

$$\begin{cases} \mathbf{W} &= \mathbf{W}^{\text{dif}} &+ & \xi \mathbf{W}^{\text{inc}}, \\ \mathbf{W} &\equiv \mathbf{W}^{\text{dif}} &\text{si} & \xi = 0, \\ \mathbf{W} &\equiv \mathbf{W}^{\text{tot}} &\text{si} & \xi = 1. \end{cases}$$

Si on note :

$$\begin{cases} \varepsilon(\mathbf{x}) &= \varepsilon_0 \varepsilon_r(\mathbf{x}), \\ \mu(\mathbf{x}) &= \mu_0 \mu_r(\mathbf{x}), \end{cases}$$

où  $\varepsilon_0$  et  $\mu_0$  désignent respectivement la permittivité et la perméabilité du vide alors on peut adimensionner le système (5) grâce au changement de variables suivant :

$$\begin{cases} \tilde{\mathbf{E}} &= \frac{\mathbf{E}}{z_0 \mathbf{H}_0}, \\ \tilde{\mathbf{H}} &= \frac{\mathbf{H}}{\mathbf{H}_0}, \\ \tilde{\omega} &= \frac{\omega}{c_0}, \end{cases}$$

où  $z_0$  est l'impédance du vide (d'une manière générale on notera par  $z$  l'impédance d'un milieu quelconque  $z = \sqrt{\frac{\mu}{\varepsilon}}$ ). Il faut aussi modifier les conditions aux bords :

– sur la *frontière métallique*  $\Gamma^m$ , on a :

$$\begin{aligned} \mathbf{n} \times \mathbf{E}^{\text{tot}} = 0 &\Leftrightarrow \mathbf{n} \times (\mathbf{E} + (1 - \xi) \mathbf{E}^{\text{inc}}) = 0 \\ &\Leftrightarrow \mathbf{n} \times \mathbf{E} = (\xi - 1) \mathbf{n} \times \mathbf{E}^{\text{inc}} \\ &\Leftrightarrow \mathbf{n} \times (z_0 \mathbf{H}_0 \tilde{\mathbf{E}}) = (\xi - 1) \mathbf{n} \times (z_0 \mathbf{H}_0 \tilde{\mathbf{E}}^{\text{inc}}). \end{aligned}$$

– sur la *frontière absorbante*  $\Gamma^a$ , on a :

$$\begin{aligned} \mathbf{n} \times \mathbf{E}^{\text{dif}} + z \mathbf{n} \times (\mathbf{n} \times \mathbf{H}^{\text{dif}}) &= 0 \\ \Leftrightarrow \mathbf{n} \times (\mathbf{E} - \xi \mathbf{E}^{\text{inc}}) + z \mathbf{n} \times (\mathbf{n} \times (\mathbf{H} - \xi \mathbf{H}^{\text{inc}})) &= 0 \\ \Leftrightarrow \mathbf{n} \times \mathbf{E} + z \mathbf{n} \times (\mathbf{n} \times \mathbf{H}) &= \xi (\mathbf{n} \times \mathbf{E}^{\text{inc}} + z \mathbf{n} \times (\mathbf{n} \times \mathbf{H}^{\text{inc}})) \\ \Leftrightarrow \mathbf{n} \times \tilde{\mathbf{E}} + z_r \mathbf{n} \times (\mathbf{n} \times \tilde{\mathbf{H}}) &= \xi (\mathbf{n} \times \tilde{\mathbf{E}}^{\text{inc}} + z_r \mathbf{n} \times (\mathbf{n} \times \tilde{\mathbf{H}}^{\text{inc}})). \end{aligned}$$

Pour alléger les notations, on abandonne les symboles  $\tilde{\cdot}$  et on fait les substitutions  $\varepsilon_r \rightarrow \varepsilon$ ,  $\mu_r \rightarrow \mu$ ,  $c_r \rightarrow c$  et  $z_r \rightarrow z$ , ce qui donne finalement après simplification :

$$\begin{cases} \mathbf{n} \times \mathbf{E} = (\xi - 1) \mathbf{n} \times \mathbf{E}^{\text{inc}} & \text{sur } \Gamma^m, \\ \mathbf{n} \times \mathbf{E} + z \mathbf{n} \times (\mathbf{n} \times \mathbf{H}) = \xi (\mathbf{n} \times \mathbf{E}^{\text{inc}} + z \mathbf{n} \times (\mathbf{n} \times \mathbf{H}^{\text{inc}})) & \text{sur } \Gamma^a. \end{cases} \quad (68)$$

Par ailleurs, si pour  $l \in (x, y, z)$  on note  $F_l(\mathbf{W}) = G_l \mathbf{W}$  avec :

$$G_l = \begin{pmatrix} 0_3 & N_l \\ -N_l & 0_3 \end{pmatrix},$$

alors le système des équations de Maxwell harmoniques s'écrit :

$$i\omega\theta\mathbf{W} + \operatorname{div}F(\mathbf{W}) + K\mathbf{W} = \mathbf{S}, \quad (69)$$

ou bien encore :

$$i\omega\theta\mathbf{W} + G_x\partial_x\mathbf{W} + G_y\partial_y\mathbf{W} + G_z\partial_z\mathbf{W} + K\mathbf{W} = \mathbf{S}. \quad (70)$$

où :

$$\mathbf{W} = \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} \quad \text{et} \quad \theta = \begin{pmatrix} \varepsilon\mathbf{I}_3 & 0_3 \\ 0_3 & \mu\mathbf{I}_3 \end{pmatrix}$$

On supposera par la suite que  $\sigma = 0$  ce qui conduit à  $K = 0$ . Rappelons aussi que si on résout les équations en champ total le second membre  $\mathbf{S}$  est nul (il est égal à  $(\mathbf{I}_6 - G_0)p\mathbf{W}^{\text{inc}}$  pour un calcul en champ diffracté). Pour la prise en compte des conditions aux limites, on procède comme en 1D c'est-à-dire que l'on associe à  $\Gamma_h$  une partition  $\mathcal{I}_b \subset \mathbb{N}$  de cellules fictives. On suppose que les paramètres  $\varepsilon$  et  $\mu$  sont constants par cellule :  $\varepsilon|_{C_j} = \varepsilon_j$  et  $\mu|_{C_j} = \mu_j$ .

#### 4.1 Méthode de volumes finis

Dans cette section, on décrit en détail la méthode de volumes finis introduite dans la section 2.4.1. Pour une fonction vectorielle  $\mathbf{W}$ , nous noterons  $\mathbf{W}_j$  sa valeur moyenne sur  $C_j$ , c'est-à-dire :

$$\mathbf{W}_j = \frac{1}{V_j} \int_{C_j} \mathbf{W} dx.$$

En intégrant l'équation (69) sur une cellule  $C_j$ , on obtient :

$$i\omega V_j \theta_j \mathbf{W}_j + \int_{C_j} \operatorname{div}F(\mathbf{W}) dx = V_j \mathbf{S}_j.$$

Grâce au théorème de Stokes :

$$\begin{aligned} i\omega V_j \theta_j \mathbf{W}_j + \int_{C_j} F(\mathbf{W}) \cdot \mathbf{n} ds &= V_j \mathbf{S}_j \\ \Leftrightarrow i\omega V_j \theta_j \mathbf{W}_j + \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} F(\mathbf{W}) \cdot \mathbf{n}_{jk} ds &= V_j \mathbf{S}_j, \end{aligned}$$

qui est équivalent au système :

$$\begin{cases} i\omega V_j \varepsilon_j \mathbf{E}_j + \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} N_{jk} \mathbf{H} ds &= -V_j \mathbf{S}_j, \\ i\omega V_j \mu_j \mathbf{H}_j - \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} N_{jk} \mathbf{E} ds &= 0. \end{cases} \quad (71)$$

où  $N_{jk} = \sum_{i=1}^3 n_{jk}^{x_i} N_i$  est basé sur des normales unitaires. On introduit les flux numériques suivants :

$$\begin{cases} \Phi_{H,jk} & \equiv \Phi_H(\mathbf{H}_j, \mathbf{H}_k, \mathbf{n}_{jk}) \approx \int_{\Sigma_{ij}} N_{ij} \mathbf{H} ds, \\ \Phi_{E,jk} & \equiv \Phi_E(\mathbf{E}_j, \mathbf{E}_k, \mathbf{n}_{jk}) \approx \int_{\Sigma_{jk}} N_{jk} \mathbf{E} ds, \end{cases} \quad (72)$$

et  $\Phi_{jk} = (\Phi_{H,jk}, -\Phi_{E,jk})^t$ . Ainsi, (71) s'écrit aussi :

$$i\omega V_j \theta_j \mathbf{W}_j + \sum_{k \in \mathcal{V}_j} \Phi_{jk} = V_j \mathbf{S}_j. \quad (73)$$

Les flux numériques (72) sont calculés de manière centrée [33] :

$$\begin{cases} \Phi_{H,jk} & = \frac{S_{jk}}{2} [N_{jk} \mathbf{H}_j + N_{jk} \mathbf{H}_k], \\ \Phi_{E,jk} & = \frac{S_{jk}}{2} [N_{jk} \mathbf{E}_j + N_{jk} \mathbf{E}_k]. \end{cases} \quad (74)$$

On introduit  $M_{jk}^j = M_{jk} \theta_j^{-1}$  avec  $M_{jk} = \begin{pmatrix} 0_3 & N_{jk} \\ -N_{jk} & 0_3 \end{pmatrix}$ .

En annexe 7.2, on montre un certain nombre de propriétés assez simples relatives à la matrice  $N_{jk}$ . Il en découle bien entendu certaines propriétés pour  $M_{ij}^i$  :

- $M_{kj}^j = M_{kj} \theta_j^{-1} = -M_{jk} \theta_j^{-1} = -M_{jk}^j$ .
- $\sum_{k \in \mathcal{V}_j} S_{jk} M_{jk}^j = (\sum_{k \in \mathcal{V}_j} S_{jk} M_{jk}) \theta_j^{-1} = 0$ .

Enfin, on définit pour  $j \in \mathcal{I}_a$ , la matrice d'impédance  $Z_j = c_j \theta_j = \begin{pmatrix} z_j^{-1} \mathbf{I}_3 & 0_3 \\ 0_3 & z_j \mathbf{I}_3 \end{pmatrix}$ . Pour revenir à notre schéma, nous devons donc étudier :

$$i\omega V_j \theta_j \mathbf{W}_j + \sum_{k \in \mathcal{V}_j} S_{jk} \mathbf{F}_{jk} = V_j \mathbf{S}_j, \quad (75)$$

avec  $\mathbf{F}_{jk} = M_{jk} \left( \frac{\mathbf{W}_j + \mathbf{W}_k}{2} \right)$ , pour  $j \in \mathcal{I}$  et  $k \in \mathcal{V}_j$ , sachant que  $\Phi_{jk} = S_{jk} \mathbf{F}_{jk}$ .

#### 4.1.1 Traitement des conditions aux limites

La démarche est strictement identique à celle suivie dans le cas 1D. Dans la mesure où les cellules  $(C_j)_{j \in \mathcal{I}_b}$  sont fictives, nous identifierons leurs paramètres à ceux de leurs cellules voisines réelles, si bien que pour  $k \in \mathcal{V}_j^m \cup \mathcal{V}_j^a$ ,  $(\varepsilon_k, \mu_k) = (\varepsilon_{|\Sigma_{ij}}, \mu_{|\Sigma_{ij}}) = (\varepsilon_j, \mu_j)$ , d'où

$M_{jk}^k = M_{jk}^j$ . Il nous faut également définir les états fictifs  $\mathbf{W}_k$  de sorte que les conditions limites soient respectées faiblement. En particulier, si nous prenons :

$$\forall (j, k) \in \mathcal{I} \times \mathcal{V}_j^a : \mathbf{W}_k = \begin{pmatrix} 0_3 & z_j N_{jk} \\ -z_j^{-1} N_{jk} & 0_3 \end{pmatrix} \mathbf{W}_j + \xi \begin{pmatrix} \mathbf{I}_3 & -z_j N_{jk} \\ z_j^{-1} N_{jk} & \mathbf{I}_3 \end{pmatrix} \mathbf{W}^{\text{inc}}, \quad (76)$$

$$\forall (j, k) \in \mathcal{I} \times \mathcal{V}_j^m : \mathbf{W}_k = \begin{pmatrix} -\mathbf{I}_3 & 0_3 \\ 0_3 & \mathbf{I}_3 \end{pmatrix} \mathbf{W}_j + 2(\xi - 1) \begin{pmatrix} \mathbf{I}_3 & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \mathbf{W}^{\text{inc}}, \quad (77)$$

alors le champ  $\frac{\mathbf{W}_j + \mathbf{W}_k}{2} \approx W_{|\Sigma_{jk}}$  satisfiera faiblement les conditions limites.

Comme dans le cas 1D (voir la section 3.1.2), la prise en compte de la condition aux limites imposée sur la frontière absorbante nécessite la diagonalisation de la matrice de flux associée à la forme conservative du système d'équations de Maxwell (4). Dans le cas présent, il s'agit de la matrice  $M_{jk}^j = M_{jk} \theta_j^{-1}$ . La diagonalisation de cette matrice est décrite dans l'annexe 7.3.

On utilise les conditions énoncées en annexe 7.4 équivalentes aux conditions limites.

**Frontières absorbantes :**  $\forall (j, k) \in \mathcal{I} \times \mathcal{V}_j^a :$

$$\begin{aligned} (M_{jk}^j)^{-\theta_j} \frac{\mathbf{W}_j + \mathbf{W}_k}{2} &= \frac{1}{2} (M_{jk}^j)^{-} \begin{pmatrix} \varepsilon_j & 0 \\ 0 & \mu_j \end{pmatrix} \begin{pmatrix} \mathbf{I}_3 & z_j^{-1} N_{jk} \\ -z_j N_{jk} & \mathbf{I}_3 \end{pmatrix} \mathbf{W}_j \\ &+ \frac{\xi}{2} (M_{jk}^j)^{-} \begin{pmatrix} \varepsilon_j & 0 \\ 0 & \mu_j \end{pmatrix} \begin{pmatrix} \mathbf{I}_3 & -z_j N_{jk} \\ z_j^{-1} N_{jk} & \mathbf{I}_3 \end{pmatrix} \mathbf{W}^{\text{inc}} \\ &= \frac{c_j}{4} \begin{pmatrix} N_{jk}^2 & z_j^{-1} N_{jk} \\ -z_j N_{jk} & N_{jk}^2 \end{pmatrix} \begin{pmatrix} \varepsilon_j & \frac{N_{jk}}{c_j} \\ -\frac{N_{jk}}{c_j} & \mu_j \end{pmatrix} \mathbf{W}_j \\ &+ \xi \frac{c_j}{4} \begin{pmatrix} N_{jk}^2 & z_j^{-1} N_{jk} \\ -z_j N_{jk} & N_{jk}^2 \end{pmatrix} \begin{pmatrix} \varepsilon_j & -\frac{N_{jk}}{c_j} \\ \frac{N_{jk}}{c_j} & \mu_j \end{pmatrix} \mathbf{W}^{\text{inc}} \\ &= \xi \frac{c_j}{2} \begin{pmatrix} N_{jk}^2 \varepsilon_j & \frac{N_{jk}}{c_j} \\ -\frac{N_{jk}}{c_j} & N_{jk}^2 \mu_j \end{pmatrix} \mathbf{W}^{\text{inc}} = \xi (M_{jk}^i)^{-\theta_j} \mathbf{W}^{\text{inc}}. \end{aligned}$$

**Frontières métalliques :**  $\forall (j, k) \in \mathcal{I} \times \mathcal{V}_j^m$  avec  $\Upsilon_{jk} = \begin{pmatrix} -N_{jk} & 0_3 \\ 0_3 & 0_3 \end{pmatrix} :$



$$\begin{aligned}
\Upsilon_{jk} \frac{\mathbf{W}_j + \mathbf{W}_k}{2} &= \Upsilon_{jk} \begin{pmatrix} 0_3 & 0_3 \\ 0_3 & \mathbf{I}_3 \end{pmatrix} \mathbf{W}_j + (1 - \xi) \Upsilon_{jk} \begin{pmatrix} -\mathbf{I}_3 & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \mathbf{W}^{\text{inc}} \\
&= \begin{pmatrix} -N_{jk} & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \begin{pmatrix} 0_3 & 0_3 \\ 0_3 & \mathbf{I}_3 \end{pmatrix} \mathbf{W}_j \\
&+ (1 - \xi) \begin{pmatrix} -N_{jk} & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \begin{pmatrix} -\mathbf{I}_3 & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \mathbf{W}^{\text{inc}} \\
&= (1 - \xi) \begin{pmatrix} N_{jk} & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \mathbf{W}^{\text{inc}} = (\xi - 1) \Upsilon_{jk} \mathbf{W}^{\text{inc}}.
\end{aligned}$$

Nous obtenons des résultats qui confirment que ces conditions sont vérifiées faiblement. On peut maintenant calculer le flux correspondant à chaque condition aux limites.

**Flux aux frontières absorbantes :**  $\forall (j, k) \in \mathcal{I} \times \mathcal{V}_j^a$  :

$$\begin{aligned}
\mathbf{F}_{jk} &= \frac{1}{2} (M_{jk} \mathbf{W}_j + M_{jk} \mathbf{W}_k) = \frac{1}{2} M_{jk} (\mathbf{W}_j + \mathbf{W}_k) \\
&= \frac{1}{2} M_{jk} \left[ \begin{pmatrix} \mathbf{I}_3 & z_j N_{jk} \\ -z_j^{-1} N_{jk} & \mathbf{I}_3 \end{pmatrix} \mathbf{W}_j + \xi \begin{pmatrix} \mathbf{I}_3 & -z_j N_{jk} \\ z_j^{-1} N_{jk} & \mathbf{I}_3 \end{pmatrix} \mathbf{W}^{\text{inc}} \right] \\
&= \frac{1}{2} \left[ \begin{pmatrix} -z_j^{-1} N_{jk}^2 & N_{jk} \\ -N_{jk} & -z_j N_{jk}^2 \end{pmatrix} \mathbf{W}_j + \xi \begin{pmatrix} z_j^{-1} N_{jk}^2 & N_{jk} \\ -N_{jk} & z_j N_{jk}^2 \end{pmatrix} \mathbf{W}^{\text{inc}} \right] \\
&= \frac{c_j}{2} \left[ \begin{pmatrix} -\varepsilon_j N_{jk}^2 & z_j^{-1} \mu_j N_{jk} \\ -z_j \varepsilon_j N_{jk} & -\mu_j N_{jk}^2 \end{pmatrix} \mathbf{W}_j + \xi \begin{pmatrix} \varepsilon_j N_{jk}^2 & z_j^{-1} \mu_j N_{jk} \\ -z_j \varepsilon_j N_{jk} & \mu_j N_{jk}^2 \end{pmatrix} \mathbf{W}^{\text{inc}} \right] \\
&= (M_{jk}^j)^+ \theta_j \mathbf{W}_j + \xi (M_{jk}^j)^- \theta_j \mathbf{W}^{\text{inc}}.
\end{aligned}$$

Comme pour le flux absorbant monodimensionnel, notre schéma à flux centrés muni de la condition (76) se ramène à un schéma à flux décentré de type Steger et Warming [37]. Les expressions des matrices  $(M_{jk}^j)^+$  et  $(M_{jk}^j)^-$  sont rappelées en annexe 7.3.

**Flux aux frontières métalliques :**  $\forall (j, k) \in \mathcal{I} \times \mathcal{V}_j^m$  :

$$\begin{aligned}
\mathbf{F}_{jk} &= \frac{1}{2}(M_{jk}\mathbf{W}_j + M_{jk}\mathbf{W}_k) = \frac{1}{2}M_{jk}(\mathbf{W}_j + \mathbf{W}_k) \\
&= \frac{1}{2}M_{jk} \left[ 2 \begin{pmatrix} 0_3 & 0_3 \\ 0_3 & \mathbf{I}_3 \end{pmatrix} \mathbf{W}_j + 2(1-\xi) \begin{pmatrix} -\mathbf{I}_3 & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \mathbf{W}^{\text{inc}} \right] \\
&= \begin{pmatrix} 0_3 & N_{jk} \\ -N_{jk} & 0_3 \end{pmatrix} \begin{pmatrix} 0_3 & 0_3 \\ 0_3 & \mathbf{I}_3 \end{pmatrix} \mathbf{W}_j + (1-\xi) \begin{pmatrix} 0_3 & N_{jk} \\ -N_{jk} & 0_3 \end{pmatrix} \begin{pmatrix} -\mathbf{I}_3 & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \mathbf{W}^{\text{inc}} \\
&= \begin{pmatrix} 0_3 & N_{jk} \\ 0_3 & 0_3 \end{pmatrix} \mathbf{W}_j + (1-\xi) \begin{pmatrix} 0_3 & 0_3 \\ N_{jk} & 0_3 \end{pmatrix} \mathbf{W}^{\text{inc}}.
\end{aligned}$$

Par conséquent, si on pose  $Q_{jk}^m = \begin{pmatrix} 0_3 & N_{jk} \\ 0_3 & 0_3 \end{pmatrix}$ , alors :

$$\begin{aligned}
\forall (j, k) \in \mathcal{I} \times \mathcal{V}_j^a : \quad \mathbf{F}_{jk}^a &= (M_{jk}^j)^+ \theta_j \mathbf{W}_j + \xi (M_{jk}^j)^- \theta_j \mathbf{W}^{\text{inc}}, \\
\forall (j, k) \in \mathcal{I} \times \mathcal{V}_j^m : \quad \mathbf{F}_{jk}^m &= Q_{jk}^m \mathbf{W}_j + (\xi - 1)(Q_{jk}^m)^t \mathbf{W}^{\text{inc}}.
\end{aligned} \tag{78}$$

sachant que  $Q_{jk}^m$  est anti-symétrique.

#### 4.1.2 Formulation matricielle

Maintenant que nous avons exprimé le flux élémentaire associé à chaque type de face du maillage  $\mathcal{T}_h$ , nous allons en écrire le bilan de manière à formuler un système matriciel similaire à celui que nous avons trouvé dans le cadre monodimensionnel. Plus précisément, pour une cellule  $\{C_j\}_{j \in \mathcal{I}}$ , on obtient :

$$\begin{aligned}
\sum_{k \in \mathcal{V}_j} S_{jk} \mathbf{F}_{jk} &= \sum_{k \in \mathcal{V}_j^d} S_{jk} \mathbf{F}_{jk} + \sum_{k \in \mathcal{V}_j^m} S_{jk} \mathbf{F}_{jk}^m + \sum_{k \in \mathcal{V}_j^a} S_{jk} \mathbf{F}_{jk}^a \\
&= \frac{1}{2} \sum_{k \in \mathcal{V}_j^d} S_{jk} (M_{jk}^j \mathbf{W}_j + M_{jk}^k \mathbf{W}_k) \\
&\quad + \sum_{k \in \mathcal{V}_j^m} S_{jk} \left( Q_{jk}^m \mathbf{W}_j + (\xi - 1)(Q_{jk}^m)^t \mathbf{W}^{\text{inc}} \right) \\
&\quad + \sum_{k \in \mathcal{V}_j^a} S_{jk} \left( (M_{jk}^j)^+ \theta_j \mathbf{W}_j + \xi (M_{jk}^j)^- \theta_j \mathbf{W}^{\text{inc}} \right)
\end{aligned}$$

Comme dans le cas 1D, les termes qui contiennent le champ incident seront assimilés à des termes sources. Nous pouvons donc les regrouper en un seul et même terme que nous noterons de la façon suivante :

$$\gamma_j^{\text{inc}} = \left( \xi \sum_{k \in \mathcal{V}_j^a} S_{jk} (M_{jk}^j)^- \theta_j + (\xi - 1) \sum_{k \in \mathcal{V}_j^m} S_{jk} (Q_{jk}^m)^t \right) \mathbf{w}^{\text{inc}}.$$

On obtient alors :

$$\begin{aligned} \sum_{k \in \mathcal{V}_j} S_{jk} \mathbf{F}_{jk} &= \frac{1}{2} \sum_{k \in \mathcal{V}_j^d} S_{jk} M_{jk} \mathbf{W}_j + \sum_{k \in \mathcal{V}_j^m} S_{jk} Q_{jk}^m \mathbf{W}_j \\ &+ \sum_{k \in \mathcal{V}_j^a} S_{jk} (M_{jk}^j)^+ \theta_j \mathbf{W}_j + \frac{1}{2} \sum_{k \in \mathcal{V}_j^d} S_{jk} M_{jk} \mathbf{W}_k + \gamma_j^{\text{inc}} \end{aligned}$$

ce qui nous conduit à :

$$\begin{aligned} \sum_{k \in \mathcal{V}_j} S_{jk} \mathbf{F}_{jk} &= \frac{1}{2} \left( \sum_{k \in \mathcal{V}_j^d} S_{jk} M_{jk} + \sum_{k \in \mathcal{V}_j^m} 2S_{jk} Q_{jk}^m + \sum_{k \in \mathcal{V}_j^a} 2S_{jk} (M_{jk}^j)^+ \theta_j \right) \mathbf{W}_j \\ &+ \frac{1}{2} \sum_{k \in \mathcal{V}_j^d} S_{jk} M_{jk} \mathbf{W}_k + \gamma_j^{\text{inc}}. \end{aligned} \quad (79)$$

La deuxième assertion de la proposition 6 de l'annexe 7.2 donne :

$$\begin{aligned} \sum_{k \in \mathcal{V}_j} S_{jk} M_{jk} = 0 &\Rightarrow \sum_{k \in \mathcal{V}_j^d} S_{jk} M_{jk} + \sum_{k \in \mathcal{V}_j^m} S_{jk} M_{jk} + \sum_{k \in \mathcal{V}_j^a} S_{jk} M_{jk} = 0 \\ &\Rightarrow \sum_{k \in \mathcal{V}_j^d} S_{jk} M_{jk} = - \sum_{k \in \mathcal{V}_j^m} S_{jk} M_{jk} - \sum_{k \in \mathcal{V}_j^a} S_{jk} M_{jk}. \end{aligned} \quad (80)$$

En injectant (80) dans (79), nous trouvons :

$$\begin{aligned} \sum_{k \in \mathcal{V}_j} S_{jk} \mathbf{F}_{jk} &= \frac{1}{2} \left( \sum_{k \in \mathcal{V}_j^m} S_{jk} (2Q_{jk}^m - M_{jk}) + \sum_{k \in \mathcal{V}_j^a} S_{jk} (2(M_{jk}^j)^+ \theta_j - M_{jk}) \right) \mathbf{W}_j \\ &+ \frac{1}{2} \sum_{k \in \mathcal{V}_j^d} S_{jk} M_{jk} \mathbf{W}_k + \gamma_j^{\text{inc}}, \end{aligned}$$

Or des calculs simples donnent :

$$\begin{aligned} 2Q_{jk}^m - M_{jk} &= \begin{pmatrix} 0_3 & N_{jk} \\ N_{jk} & 0_3 \end{pmatrix}, \\ 2(M_{jk}^j)^+ \theta_j - M_{jk} &= \begin{pmatrix} -z_j^{-1} N_{jk}^2 & 0_3 \\ 0_3 & -z_j N_{jk}^2 \end{pmatrix}. \end{aligned}$$

Si on pose  $P_{jk}^m = \begin{pmatrix} 0_3 & N_{jk} \\ N_{jk} & 0_3 \end{pmatrix}$  et  $P_{jk}^a = \begin{pmatrix} -z_j^{-1} N_{jk}^2 & 0_3 \\ 0_3 & -z_j N_{jk}^2 \end{pmatrix}$ , alors :

$$\sum_{k \in \mathcal{V}_j} S_{jk} \mathbf{F}_{jk} = \frac{1}{2} \left( \sum_{k \in \mathcal{V}_j^m} S_{jk} P_{jk}^m + \sum_{k \in \mathcal{V}_j^a} S_{jk} P_{jk}^a \right) \mathbf{W}_j + \frac{1}{2} \sum_{k \in \mathcal{V}_j^d} S_{jk} M_{jk} \mathbf{W}_k + \gamma_j^{\text{inc}}. \quad (81)$$

La dernière étape consiste à injecter (81) dans (75). On obtient :

$$i\omega V_j \mathbf{W}_j + \frac{1}{2} \left( \sum_{k \in \mathcal{V}_j^m} S_{jk} P_{jk}^m + \sum_{k \in \mathcal{V}_j^a} S_{jk} P_{jk}^a \right) \mathbf{W}_j + \frac{1}{2} \sum_{k \in \mathcal{V}_j^d} S_{jk} M_{jk} \mathbf{W}_k + \gamma_j^{\text{inc}} = V_j \mathbf{S}_j,$$

pour aboutir au schéma suivant : pour tout  $j \in \mathcal{I}$  :

$$\left( 2i\omega V_j \theta_j + \sum_{k \in \mathcal{V}_j^m} S_{jk} P_{jk}^m + \sum_{k \in \mathcal{V}_j^a} S_{jk} P_{jk}^a \right) \mathbf{W}_j + \sum_{k \in \mathcal{V}_j^d} S_{jk} M_{jk} \mathbf{W}_k = 2(V_j \mathbf{S}_j - \gamma_j^{\text{inc}}). \quad (82)$$

Il s'agit maintenant d'écrire une formulation matricielle du schéma (82). Ce faisant, nous voulons montrer que la matrice globale du système possède une structure identique à celle du cas 1D. On désigne par  $\mathbf{b} \in \mathbb{C}^{6 \times N}$  le second membre et par  $\mathcal{A} \in \mathbf{M}_{6 \times N}(\mathbb{C})$  la matrice globale résultant de l'assemblage du schéma (82).

Pour  $\mathbf{W} \in \mathbb{C}^{6 \times N}$ , on trouve grâce au bilan des flux pour la cellule  $C_j$  que (82) s'écrit sous la forme  $(\mathcal{A}\mathbf{W})_j = b_j$  avec  $\mathcal{A} = i\mathcal{D} + \mathcal{Z} + \mathcal{R}$  et :

$$\begin{cases} (\mathcal{D}\mathbf{W})_j = 2\omega V_j \theta_j \mathbf{W}_j, \\ (\mathcal{Z}\mathbf{W})_j = \sum_{k \in \mathcal{V}_j^a} S_{jk} P_{jk}^a \mathbf{W}_j, \\ (\mathcal{R}\mathbf{W})_j = \sum_{k \in \mathcal{V}_j^m} S_{jk} P_{jk}^m \mathbf{W}_j + \sum_{k \in \mathcal{V}_j^d} S_{jk} M_{jk} \mathbf{W}_k, \\ b_j = 2(V_j \mathbf{S}_j - \gamma_j^{\text{inc}}). \end{cases}$$

**Proposition 3** Les matrices  $\mathcal{Z}$ ,  $\mathcal{R}$  et  $\mathcal{D}$  ont les propriétés suivantes :

- (i)  $\mathcal{D}$  est réelle symétrique définie positive,
- (ii)  $\mathcal{Z}$  est réelle symétrique et positive,
- (iii)  $\mathcal{R}$  est réelle antisymétrique.

**Preuve**

(i)  $\mathcal{D}$  est clairement diagonale par blocs, avec  $\mathcal{D}_{jj} = 2\omega V_j \theta_j$  qui est symétrique définie positive. On a donc que  $\mathcal{D}$  est symétrique définie positive.

(ii)  $\mathcal{Z}$  est aussi diagonale par blocs et  $\mathcal{Z}_{jj} = \sum_{k \in \mathcal{V}_j^a} S_{jk} P_{jk}^a$ . De plus,  $P_{jk}^a$  est symétrique

positive (voir proposition 13). Donc  $\sum_{j \in \mathcal{V}_j^a} S_{jk} P_{jk}^a$  est symétrique positive. On en déduit que

$\mathcal{Z}$  est symétrique et positive.

(iii) D'après son développement :

$$\forall j \in \mathcal{I}, \mathcal{R}_{jj} = \sum_{k \in \mathcal{V}_j^m} S_{jk} P_{jk}^m \text{ et } \mathcal{R}_{jk} = \begin{cases} S_{jk} M_{jk} & \text{pour } k \in \mathcal{V}_j^d, \\ 0 & \text{pour } k \in \mathcal{I} \setminus (\{j\} \cup \mathcal{V}_j^d). \end{cases}$$

Nous avons prouvé en proposition 13 de l'annexe 7.6 que  $P_{jk}^m$  et  $M_{jk}$  sont respectivement antisymétrique et symétrique, on en déduit que :

$$\forall j \in \mathcal{I}, \mathcal{R}_{jj} \text{ est antisymétrique et } \forall (j, k) \in \mathcal{I}^2, j \neq k, \mathcal{R}_{jk} \text{ est symétrique}$$

D'autre part, nous avons aussi que  $\mathcal{R}_{ji} = -\mathcal{R}_{jk}$  puisque  $M_{ji} = -M_{jk}$ . On est en mesure d'appliquer la proposition 11 de l'annexe 7.6, et donc de conclure que  $\mathcal{R}$  est antisymétrique. ■

En résumé, nous avons une matrice ayant une structure identique à la matrice obtenue en 1D.

**4.1.3 Étude de la dispersion numérique**

Dans ce paragraphe, nous allons étudier la dispersion numérique de notre schéma. L'étude de la dispersion d'un schéma permet de connaître sa capacité à propager correctement une onde. En effet, l'ordre de précision de la relation de dispersion discrète est souvent mise en avant pour évaluer la précision d'un schéma numérique. L'étude de dispersion consiste à s'intéresser à des solutions particulières des équations de Maxwell, appelées ondes planes, qui sont de la forme :

$$\mathbf{U}(x, y, z) = \mathbf{U}_0 e^{-i\mathbf{k} \cdot (x, y, z)^t}.$$

En injectant cette onde dans les EM continues, on trouve la relation de dispersion *exacte*  $\omega_{ex}^2 = c^2 |\mathbf{k}|^2$ .

En considérant une grille cartésienne dont les cellules sont de taille  $\Delta x \times \Delta y \times \Delta z$  et dont les caractéristiques du milieu  $\varepsilon$  et  $\mu$  sont constantes, nous pouvons rechercher les solutions particulières de notre schéma, de type ondes planes numériques, qui sont de la

forme suivante : pour chaque cellule  $C_{j,l,m}$ , dont le centre de gravité a pour coordonnées  $(j\Delta x, l\Delta y, m\Delta z)$  :

$$\mathbf{U}_{j,l,m} = \mathbf{U}_0 e^{-i(jk_x \Delta x + lk_y \Delta y + mk_z \Delta z)},$$

ce qui nous amènera à connaître la relation de dispersion discrète entre  $\omega$ ,  $\mathbf{k}$  le nombre d'onde et de  $c$ . Sous les hypothèses précédentes, notre schéma (82) s'écrit :

$$\begin{aligned} i\omega \Delta x \Delta y \Delta z \theta_{j,l,m} \mathbf{X}_{j,l,m} &+ \Delta y \Delta z \frac{M_{j+1,l,m} \mathbf{X}_{j+1,l,m} + M_{j-1,l,m} \mathbf{X}_{j-1,l,m}}{2} \\ &+ \Delta x \Delta z \frac{M_{j,l+1,m} \mathbf{X}_{j,l+1,m} + M_{j,l-1,m} \mathbf{X}_{j,l-1,m}}{2} \\ &+ \Delta x \Delta y \frac{M_{j,l,m+1} \mathbf{X}_{j,l,m+1} + M_{j,l,m-1} \mathbf{X}_{j,l,m-1}}{2} = 0, \end{aligned}$$

et conduit au schéma différences finies centrées suivant :

$$i\omega \begin{pmatrix} E_{j,l,m}^x \\ E_{j,l,m}^y \\ E_{j,l,m}^z \end{pmatrix} = \frac{1}{\mu} \begin{pmatrix} \frac{H_{j,l,m+1}^y - H_{j,l,m-1}^y}{2\Delta z} - \frac{H_{j,l+1,m}^z - H_{j,l-1,m}^z}{2\Delta y} \\ \frac{H_{j+1,l,m}^z - H_{j-1,l,m}^z}{2\Delta x} - \frac{H_{j,l,m+1}^x - H_{j,l,m-1}^x}{2\Delta z} \\ \frac{H_{j,l+1,m}^x - H_{j,l-1,m}^x}{2\Delta y} - \frac{H_{j+1,l,m}^y - H_{j-1,l,m}^y}{2\Delta x} \end{pmatrix} \quad (83)$$

et :

$$i\omega \begin{pmatrix} H_{j,l,m}^x \\ H_{j,l,m}^y \\ H_{j,l,m}^z \end{pmatrix} = -\frac{1}{\varepsilon} \begin{pmatrix} \frac{E_{j,l,m+1}^y - E_{j,l,m-1}^y}{2\Delta z} - \frac{E_{j,l+1,m}^z - E_{j,l-1,m}^z}{2\Delta y} \\ \frac{E_{j+1,l,m}^z - E_{j-1,l,m}^z}{2\Delta x} - \frac{E_{j,l,m+1}^x - E_{j,l,m-1}^x}{2\Delta z} \\ \frac{E_{j,l+1,m}^x - E_{j,l-1,m}^x}{2\Delta y} - \frac{E_{j+1,l,m}^y - E_{j-1,l,m}^y}{2\Delta x} \end{pmatrix}. \quad (84)$$

En injectant (84) dans (83) et utilisant de nouveau (84), et en supposant vérifié au niveau discret la relation  $\nabla(\text{div } \mathbf{E}) = 0^3$ , on obtient :

$$\begin{aligned} \omega^2 \mathbf{E}_{j,l,m} = c^2 \left( \frac{-\mathbf{E}_{j-2,l,m} + 2\mathbf{E}_{j,l,m} - \mathbf{E}_{j+2,l,m}}{4\Delta x^2} + \frac{-\mathbf{E}_{j,l-2,m} + 2\mathbf{E}_{j,l,m} - \mathbf{E}_{j,l+2,m}}{4\Delta y^2} \right. \\ \left. + \frac{-\mathbf{E}_{j,l,m-2} + 2\mathbf{E}_{j,l,m} - \mathbf{E}_{j,l,m+2}}{4\Delta z^2} \right) \end{aligned} \quad (85)$$

Il se trouve que (85) correspond exactement à une discrétisation de l'équation de Helmholtz par un schéma différences centrées. En effet, si on part des équations de Maxwell en domaine fréquentiel, on obtient sous l'hypothèse que  $\varepsilon$  et  $\mu$  sont constantes :

<sup>3</sup>En l'absence de densité de charge volumique la loi de Gauss donne  $\text{div } \mathbf{E} = 0$ .

$$\begin{aligned}
\begin{cases} i\omega \mathbf{E} - \nabla \times (\mu^{-1} \mathbf{H}) = 0 \\ i\omega \mathbf{H} + \nabla \times (\varepsilon^{-1} \mathbf{E}) = 0 \end{cases} &\Leftrightarrow \begin{cases} i\omega \mathbf{E} - \mu^{-1} \text{rot} \mathbf{H} = 0 \\ \mathbf{H} = \frac{i}{\varepsilon\omega} \text{rot} \mathbf{E} \end{cases} \\
&\Leftrightarrow \omega^2 \mathbf{E} + c^2 (\Delta \mathbf{E} - \nabla(\text{div} \mathbf{E})) = 0, \\
&\Leftrightarrow \omega^2 \mathbf{E} + c^2 \Delta \mathbf{E} = 0 \quad (\text{sous l'hypothèse } \nabla(\text{div} \mathbf{E}) = 0),
\end{aligned}$$

dont la discrétisation en différences centrées aboutit à (85), ce qui montre le lien très étroit entre l'équation de Helmholtz et les équations de Maxwell en domaine fréquentiel. Nous allons maintenant évaluer les termes du second membre de (85) en y injectant l'onde plane numérique définie auparavant. On trouve :

$$\begin{aligned}
\frac{-\mathbf{E}_{j-2,l,m} + 2\mathbf{E}_{j,l,m} - \mathbf{E}_{j+2,l,m}}{4\Delta x^2} &= \frac{-e^{j2k_x\Delta x} + 2 - e^{-j2k_x\Delta x}}{4\Delta x^2} \mathbf{E}_{j,l,m} \\
&= \frac{2 - 2\cos(2k_x\Delta x)}{4\Delta x^2} \mathbf{E}_{j,l,m} \\
&= \frac{\sin^2(k_x\Delta x)}{\Delta x^2} \mathbf{E}_{j,l,m}.
\end{aligned}$$

On en déduit donc que, pour une solution discrète non identiquement nulle,  $\omega$  et  $\mathbf{k}$  sont liés par la relation de dispersion discrète suivante :

$$\frac{\omega^2}{c^2} = \frac{\sin^2(k_x\Delta x)}{\Delta x^2} + \frac{\sin^2(k_y\Delta y)}{\Delta y^2} + \frac{\sin^2(k_z\Delta z)}{\Delta z^2}.$$

Cette relation signifie que les ondes planes ne se propagent pas exactement avec la bonne pulsation ou la bonne vitesse. On remarque néanmoins qu'on retrouve la relation de dispersion exacte quand les trois pas d'espace  $\Delta x$ ,  $\Delta y$  et  $\Delta z$  tendent vers 0.

En effectuant un développement limité en  $\Delta x$ ,  $\Delta y$  et  $\Delta z$  en 0 d'ordre 5 dans (85), on trouve, en notant  $h = \max(\Delta x, \Delta y, \Delta z)$ , la relation de dispersion approchée :

$$\frac{\omega^2}{\omega_{ex}^2} = 1 - \frac{k_x^4\Delta x^2 + k_y^4\Delta y^2 + k_z^4\Delta z^2}{3|k|^2} + O(|k|^4 h^4). \quad (86)$$

Il est intéressant de comparer cette relation à celle obtenue dans le domaine temporel. En effet, en se référant à [35], l'erreur de dispersion dans le domaine temporel pour ce même schéma en espace et pour un schéma explicite de type saute-mouton pour l'intégration en temps est donnée par :

$$\frac{\omega^2}{\omega_{ex}^2} \left( 1 - \frac{\omega_{ex}^2 \Delta t^2}{12} + O(\omega_{ex}^4 \Delta t^4) \right) = 1 - \frac{k_x^4\Delta x^2 + k_y^4\Delta y^2 + k_z^4\Delta z^2}{3|k|^2} + O(|k|^4 h^4).$$

Nous remarquons que notre relation de dispersion est la même que la relation en temporel lorsque le pas de temps tend vers 0. Mais nous constatons aussi que notre schéma est également plus dispersif que le schéma en temporel, cette différence diminuant lorsque l'on prend un pas de temps de plus en plus petit.

## 4.2 Méthode de type Galerkin discontinu

Dans cette section, on décrit en détail la méthode de type Galerkin discontinu introduite dans la section 2.4.2. Nous considérons de nouveau l'équation :

$$i\omega\theta\mathbf{W} + \operatorname{div}F(\mathbf{W}) = \mathbf{S}. \quad (87)$$

On multiplie (87) par une fonction test scalaire  $\varphi_j$  et on intègre sur une cellule  $C_j$  :

$$\begin{aligned} \int_{C_j} (i\omega\theta\mathbf{W} + \operatorname{div}F(\mathbf{W}))\varphi_j dx &= \int_{C_j} \mathbf{S}\varphi_j dx \\ \Leftrightarrow i\omega\theta_j \int_{C_j} \mathbf{W}\varphi_j dx + \int_{C_j} (\operatorname{div}F(\mathbf{W}))\varphi_j dx &= \int_{C_j} \mathbf{S}\varphi_j dx. \end{aligned}$$

En appliquant le théorème de Green-Riemann, on obtient :

$$i\omega\theta_j \int_{C_j} \mathbf{W}\varphi_j dx - \int_{C_j} F(\mathbf{W}) \cdot \nabla\varphi_j dx + \int_{\partial C_j} (F(\mathbf{W}) \cdot \mathbf{n}) \varphi_j ds = \int_{C_j} \mathbf{S}\varphi_j dx. \quad (88)$$

Comme dans le cas 1D, on pose  $\phi_j = (\varphi_{j1}, \varphi_{j2}, \dots, \varphi_{jd_j})$  la base locale de  $\mathcal{P}_j = \mathbb{P}_m[C_j]$  et on note  $\mathbf{W}_j$  la projection de  $\mathbf{W}$  sur  $\mathcal{P}_j$  (de même,  $\mathbf{S}_j$  est la projection de  $\mathbf{S}$  sur  $\mathcal{P}_j$ ). Il s'agit maintenant d'injecter  $\mathbf{W}_j$  dans l'équation (88). Notre problème est alors de chercher  $\mathbf{W}_j \in \mathcal{P}_j$  tel que :

$$\begin{aligned} i\omega\theta_j \int_{C_j} \mathbf{W}_j\varphi_j dx &- \int_{C_j} F(\mathbf{W})_j \cdot \nabla\varphi_j dx \\ &+ \int_{\partial C_j} (F(\mathbf{W}) \cdot \mathbf{n}) \varphi_j ds = \int_{C_j} \mathbf{S}\varphi_j dx \\ \Leftrightarrow i\omega\theta_j \int_{C_j} \mathbf{W}_j\varphi_j dx &- \int_{C_j} F(\mathbf{W})_j \cdot \nabla\varphi_j dx \\ &+ \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} (F(\mathbf{W}) \cdot \mathbf{n}_{jk}) \varphi_j ds = \int_{C_j} \mathbf{S}_j\varphi_j dx. \end{aligned} \quad (89)$$

Il reste à approcher les flux sur  $C_j$  ainsi que sur  $\Sigma_{jk}$ . Nous choisissons d'approcher le flux sur  $\Sigma_{jk}$  par un schéma centré :

$$F(\mathbf{W})|_{\Sigma_{jk}} \approx \frac{F(\mathbf{W})_j + F(\mathbf{W})_k}{2}.$$



Par conséquent :

$$\begin{aligned}
\sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} (F(\mathbf{W}) \cdot \mathbf{n}_{jk}) \varphi_j ds &= \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} \left( \frac{F(\mathbf{W})_j + F(\mathbf{W})_k}{2} \cdot \mathbf{n}_{jk} \right) \varphi_j ds \\
&= \frac{1}{2} \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} (F(\mathbf{W})_j \cdot \mathbf{n}_{jk}) \varphi_j ds \\
&\quad + \frac{1}{2} \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} (F(\mathbf{W})_k \cdot \mathbf{n}_{jk}) \varphi_j ds \\
&= \frac{1}{2} \int_{\partial C_j} (F(\mathbf{W})_j \cdot \mathbf{n}_{jk}) \varphi_j ds \\
&\quad + \frac{1}{2} \int_{\partial C_j} (F(\mathbf{W})_k \cdot \mathbf{n}_{jk}) \varphi_j ds.
\end{aligned}$$

En appliquant dans l'autre sens le théorème de Green-Riemann au premier terme du membre de droite, on obtient :

$$\begin{aligned}
\sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} (F(\mathbf{W}) \cdot \mathbf{n}_{jk}) \varphi_j ds &= \frac{1}{2} \int_{C_j} ((\operatorname{div} F(\mathbf{W}))_j \varphi_j + F(\mathbf{W})_j \cdot \nabla \varphi_j) dx \\
&\quad + \frac{1}{2} \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} (F(\mathbf{W})_k \cdot \mathbf{n}_{jk}) \varphi_j ds.
\end{aligned} \tag{90}$$

En injectant (90) dans la dernière équation de (89), on aboutit à :

$$\begin{aligned}
i\omega \theta_j \int_{C_j} \mathbf{W}_j \varphi_j dx &- \int_{C_j} F(\mathbf{W})_j \cdot \nabla \varphi_j dx \\
&+ \frac{1}{2} \int_{C_j} ((\operatorname{div} F(\mathbf{W}))_j \varphi_j + F(\mathbf{W})_j \cdot \nabla \varphi_j) dx \\
&+ \frac{1}{2} \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} (F(\mathbf{W})_k \cdot \mathbf{n}_{jk}) \varphi_j ds = \int_{C_j} \mathbf{S}_j \varphi_j dx.
\end{aligned} \tag{91}$$

On pose :

$$F(\mathbf{W})_j \equiv F(\mathbf{W})|_{C_j} \approx F_j(\mathbf{W}_j).$$

Par suite, l'équation (91) devient :

$$\begin{aligned}
i\omega\theta_j \int_{C_j} \mathbf{W}_j \varphi_j dx & - \int_{C_j} F_j(\mathbf{W}_j) \cdot \nabla \varphi_j dx \\
& + \frac{1}{2} \int_{C_j} ((\operatorname{div} F_j(\mathbf{W}_j)) \varphi_j + F_j(\mathbf{W}_j) \cdot \nabla \varphi_j) dx \\
& + \frac{1}{2} \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} (F_k(\mathbf{W}_k) \cdot \mathbf{n}_{jk}) \varphi_j ds = \int_{C_j} \mathbf{S}_j \varphi_j dx.
\end{aligned} \tag{92}$$

En regroupant les termes de gradients et de divergence, (92) devient :

$$\begin{aligned}
i\omega\theta_j \int_{C_j} \mathbf{W}_j \varphi_j dx & + \frac{1}{2} \int_{C_j} ((\operatorname{div} F_j(\mathbf{W}_j)) \varphi_j - F_j(\mathbf{W}_j) \cdot \nabla \varphi_j) dx \\
& + \frac{1}{2} \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} (F_k(\mathbf{W}_k) \cdot \mathbf{n}_{jk}) \varphi_j ds = \int_{C_j} \mathbf{S}_j \varphi_j dx,
\end{aligned} \tag{93}$$

qui, après développement, donne :

$$\begin{aligned}
i\omega\theta_j \int_{C_j} \mathbf{W}_j \varphi_j dx & + \frac{1}{2} \int_{C_j} \left( \sum_{l=1}^3 \partial_{x_l} (G_{x_l})_j \mathbf{W}_j \right) \varphi_j dx - \frac{1}{2} \int_{C_j} \sum_{l=1}^3 (\partial_{x_l} \varphi_j) (G_{x_l})_j \mathbf{W}_j dx \\
& + \frac{1}{2} \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} \left( \sum_{l=1}^3 n_{jk}^{x_l} (G_{x_l})_k \mathbf{W}_k \right) \varphi_j ds = \int_{C_j} \mathbf{S}_j \varphi_j dx,
\end{aligned} \tag{94}$$

avec  $\partial_{x_l} (G_{x_l})_j \mathbf{W}_j \equiv \frac{\partial (G_{x_l})_j \mathbf{W}_j}{\partial x_l}$ ,  $\partial_{x_l} \varphi_j \equiv \frac{\partial \varphi_j}{\partial x_l}$  et  $\mathbf{n}_{jk} = (n_{jk}^{x_1}, n_{jk}^{x_2}, n_{jk}^{x_3})^t$ . L'équation (94) peut donc se réécrire :

$$\begin{aligned}
i\omega\theta_j \int_{C_j} \mathbf{W}_j \varphi_j dx & + \frac{1}{2} \int_{C_j} \left( \sum_{l=1}^3 \partial_{x_l} (G_{x_l})_j \mathbf{W}_j - \sum_{l=1}^3 (\partial_{x_l} \varphi_j) (G_{x_l})_j \mathbf{W}_j \right) \varphi_j dx \\
& + \frac{1}{2} \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} M_{jk} \mathbf{W}_k \varphi_j ds = \int_{C_j} \mathbf{S}_j \varphi_j dx,
\end{aligned} \tag{95}$$

où  $M_{jk}^k = M_{jk} \theta_k^{-1}$  si  $M_{jk} = \begin{pmatrix} 0_3 & N_{jk} \\ -N_{jk} & 0_3 \end{pmatrix}$  et  $\theta_k = \begin{pmatrix} \varepsilon_k & 0_3 \\ 0_3 & \mu_k \end{pmatrix}$ .

#### 4.2.1 Traitement des conditions aux limites

On approche les flux sur les cellules qui s'appuient sur les frontières physiques comme pour le schéma volumes finis. On utilise donc les relations (76) et (77), ainsi que les matrices qui seront utiles pour les prochains calculs :

$$Q_{jk}^m = \begin{pmatrix} 0_3 & N_{jk} \\ 0_3 & 0_3 \end{pmatrix}, \quad Z_j = \begin{pmatrix} z_j^{-1} \mathbf{I}_3 & 0_3 \\ 0_3 & z_j \mathbf{I}_3 \end{pmatrix},$$

$$P_{jk}^a = \begin{pmatrix} -z_j^{-1} N_{jk}^2 & 0_3 \\ 0_3 & -z_j N_{jk}^2 \end{pmatrix}, \quad P_{jk}^m = \begin{pmatrix} 0_3 & N_{jk} \\ N_{jk} & 0_3 \end{pmatrix}.$$

Nous pouvons alors exprimer les flux frontières. Pour tout  $j \in \mathcal{I}$  et tout  $k \in \mathcal{V}_i^a$  :

$$\begin{aligned} M_{jk} \mathbf{W}_k &= M_{jk} \left( \begin{pmatrix} 0_3 & z_j N_{jk} \\ -z_j^{-1} N_{jk} & 0_3 \end{pmatrix} \mathbf{W}_j + \xi \begin{pmatrix} \mathbf{I}_3 & -z_j N_{jk} \\ z_j^{-1} N_{jk} & \mathbf{I}_3 \end{pmatrix} \mathbf{W}^{\text{inc}} \right) \\ &= M_{jk}^2 Z_j \mathbf{W}_j + \xi M_{jk} (\mathbf{I}_6 - M_{jk} Z_j) \mathbf{W}^{\text{inc}} \\ &= M_{jk}^2 Z_j \mathbf{W}_j + \xi (M_{jk} - M_{jk}^2 Z_j) \mathbf{W}^{\text{inc}} \\ &= P_{jk}^a \mathbf{W}_j + 2\xi (M_{jk}^j)^- \theta_j \mathbf{W}^{\text{inc}}. \end{aligned} \tag{96}$$

Pour tout  $j \in \mathcal{I}$  et tout  $k \in \mathcal{V}_j^m$  :

$$\begin{aligned} M_{jk} \mathbf{W}_k &= M_{jk} \mathbf{W}_k \\ &= M_{jk} \left( \begin{pmatrix} -\mathbf{I}_3 & 0_3 \\ 0_3 & \mathbf{I}_3 \end{pmatrix} \mathbf{W}_j + 2(1 - \xi) \begin{pmatrix} -\mathbf{I}_3 & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \mathbf{W}^{\text{inc}} \right) \\ &= P_{jk}^m \mathbf{W}_j + 2(\xi - 1)(Q_{jk}^m)^t \mathbf{W}^{\text{inc}}. \end{aligned} \tag{97}$$

Maintenant, si on se sert des relations (96) et (97) dans la quantité ci-dessous :

$$\begin{aligned} \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} M_{jk} \mathbf{W}_k \varphi_j ds &= \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} M_{jk} \mathbf{W}_k \varphi_j ds + \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} M_{jk} \mathbf{W}_k \varphi_j ds \\ &+ \sum_{k \in \mathcal{V}_j^m} \int_{\Sigma_{jk}} M_{jk} \mathbf{W}_k \varphi_j ds, \end{aligned} \tag{98}$$

on trouve que les flux aux cellules voisines s'écrivent :

$$\begin{aligned} \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} M_{jk} \mathbf{W}_k \varphi_j ds &= \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} M_{jk} \mathbf{W}_k \varphi_j ds \\ &+ \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} \left( P_{jk}^a \mathbf{W}_j + 2\xi (M_{jk}^j)^- \theta_j \mathbf{W}^{\text{inc}} \right) \varphi_j ds \\ &+ \sum_{k \in \mathcal{V}_j^m} \int_{\Sigma_{jk}} \left( P_{jk}^m \mathbf{W}_j + 2(\xi - 1)(Q_{jk}^m)^t \mathbf{W}^{\text{inc}} \right) \varphi_j ds, \end{aligned}$$

soit encore :

$$\begin{aligned} \sum_{k \in \mathcal{V}_j} \int_{\Sigma_{jk}} M_{jk} \mathbf{W}_k \varphi_j ds &= \sum_{k \in \mathcal{V}_j^d} \int_{\Sigma_{jk}} M_{jk} \mathbf{W}_k \varphi_j ds + \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} P_{jk}^a \mathbf{W}_j \varphi_j ds \\ &+ \sum_{k \in \mathcal{V}_j^m} \int_{\Sigma_{jk}} P_{jk}^m \mathbf{W}_j \varphi_j ds + 2\gamma_j^{\text{inc}}, \end{aligned} \quad (99)$$

avec :

$$\gamma_j^{\text{inc}} = \xi \sum_{k \in \mathcal{V}_j^a} \int_{\Sigma_{jk}} (M_{jk}^j)^{-} \theta_j \mathbf{W}^{\text{inc}} \varphi_j ds + (\xi - 1) \sum_{k \in \mathcal{V}_j^m} \int_{\Sigma_{jk}} (Q_{jk}^m)^t \mathbf{W}^{\text{inc}} \varphi_j ds.$$

Comme dans le cas 1D, nous supposons maintenant que  $\mathbf{W}_j(\mathbf{x}) = \sum_{k=1}^{d_j} \mathbf{W}_{jk} \varphi_{jk}(\mathbf{x})$ , où  $\mathcal{P}_j = \mathbb{P}_m[C_j]$  est pris comme l'espace engendré par les fonctions scalaires polynômiales de degré au plus égal à  $m$  sur la cellule  $C_j$  et les degrés de liberté locaux sont  $\mathbf{W}_{jk} \in \mathbb{C}^6$ . Posons  $\mathbb{W}_j = (\mathbf{W}_{j1}, \mathbf{W}_{j2}, \dots, \mathbf{W}_{jd_j})$  ( $\mathbb{W}_j$  est une matrice rectangle  $6 \times d_j$ ) et  $\phi_j = (\varphi_{j1}, \varphi_{j2}, \dots, \varphi_{jd_j})$ . On a  $\mathbf{W}_j = \mathbb{W}_j \phi_j^t$ . On introduit alors les matrices suivantes :

$$\left\{ \begin{array}{ll} \Phi_j &= \int_{C_j} \phi_j^t \phi_j dx & (\Phi_j \text{ est symétrique définie positive}), \\ \Phi_j^{x_l} &= \int_{C_j} \left( \phi_j^t (\partial_{x_l} \phi_j) - (\partial_{x_l} \phi_j)^t \phi_j \right) dx & (\Phi_j^{x_l} \text{ est antisymétrique}), \\ \Phi_{jk} &= \int_{\Sigma_{jk}} \phi_j^t \phi_k ds & (\Phi_{jk} \text{ est quelconque de taille } d_j \times d_k). \end{array} \right.$$

Maintenant, en prenant pour  $\varphi_j$  successivement toutes les fonctions de base  $\varphi_{jk}$  pour  $1 \leq k \leq d_j$ , on peut reformuler l'équation (95) sous la forme,  $\forall C_j \in \mathcal{T}_h$  :

$$\begin{aligned} 2i\omega\theta_j \mathbb{W}_j \Phi_j^t + \sum_{l=1}^3 G_{x_l} \mathbb{W}_j (\Phi_j^{x_l})^t &+ \sum_{k \in \mathcal{V}_j^a} P_{jk}^a \mathbb{W}_j (\Phi_{jk})^t \\ &+ \sum_{k \in \mathcal{V}_j^m} P_{jk}^m \mathbb{W}_j (\Phi_{jk})^t \\ &+ \sum_{k \in \mathcal{V}_j^d} M_{jk} \mathbb{W}_k (\Phi_{jk})^t = \mathbf{b}_j. \end{aligned} \quad (100)$$

où :

$$\mathbf{b}_j = 2 \left( \mathbb{S}_j \Phi_j^t - 2 \left( \xi \sum_{k \in \mathcal{V}_j^a} (M_{jk}^j)^{-} \theta_j \mathbb{W}^{\text{inc}}(\Phi_{jk})^t + (\xi - 1) \sum_{k \in \mathcal{V}_j^m} (Q_{jk}^m)^t \mathbb{W}^{\text{inc}}(\Phi_{jk})^t \right) \right)$$

On constate que cette formulation se rapproche de celle établie dans le cas 1D, ce qui va nous amener à réutiliser le procédé de vectorisation qui transforme la matrice  $\mathbb{W}_j = (\mathbf{W}_{j1}, \mathbf{W}_{j2}, \dots, \mathbf{W}_{jd_j})$  en un vecteur colonne  $(\mathbf{W}_{j1}, \mathbf{W}_{j2}, \dots, \mathbf{W}_{jd_j})^t \in \mathbb{C}^{6 \times d_j}$ . Ce rangement correspond toujours à un empilement par coordonnées selon un même degré de liberté. Nous obtenons alors grâce au produit de Kronecker :

$$\begin{aligned} 2i\omega(\Phi_j \otimes \theta_j) \aleph(\mathbb{W}_j) + \sum_{l=1}^3 (\Phi_j^{x_l} \otimes G_{x_l}) \aleph(\mathbb{W}_j) &+ \sum_{k \in \mathcal{V}_j^a} (\Phi_{jk} \otimes P_{jk}^a) \aleph(\mathbb{W}_j) \\ &+ \sum_{k \in \mathcal{V}_j^m} (\Phi_{jk} \otimes P_{jk}^m) \aleph(\mathbb{W}_j) \\ &+ \sum_{k \in \mathcal{V}_j^d} (\Phi_{jk} \otimes M_{jk}) \aleph(\mathbb{W}_k) = \aleph(\mathbf{b}_j), \end{aligned} \quad (101)$$

qui, après factorisation, donne :

$$\begin{aligned} \left[ 2i\omega(\Phi_j \otimes \theta_j) + \sum_{l=1}^3 (\Phi_j^{x_l} \otimes G_{x_l}) \right. &+ \sum_{k \in \mathcal{V}_j^a} (\Phi_{jk} \otimes P_{jk}^a) \\ &+ \sum_{k \in \mathcal{V}_j^m} (\Phi_{jk} \otimes P_{jk}^m) \left. \right] \aleph(\mathbb{W}_j) \\ &+ \sum_{k \in \mathcal{V}_j^d} (\Phi_{jk} \otimes M_{jk}) \aleph(\mathbb{W}_k) = \aleph(\mathbf{b}_j). \end{aligned} \quad (102)$$

On voit donc que l'on obtient une formulation matricielle dont la matrice diagonale est constituée de blocs de taille  $d_j^2$ , et dont les matrices extradiagonales sont constituées de blocs de taille  $d_j \times d_k$ . On peut aussi ordonner les inconnues d'une autre manière c'est-à-dire les ranger en vecteur où cette fois-ci ce sont les lignes de la matrice  $\mathbb{W}$  qui ont été empilées plutôt que ses colonnes. C'est un rangement par degrés de liberté successifs pour une même coordonnée, ce qui est en quelque sorte le rangement alternatif au précédent. En particulier, ce vecteur correspond à  $\aleph(\mathbb{W}^t)$ . Le schéma relatif à cet empilement s'écrit :

$$\begin{aligned}
\left[ 2i\omega(\theta_j \otimes \Phi_j) + \sum_{l=1}^3 (G_{x_l} \otimes \Phi_j^{x_l}) + \sum_{k \in \mathcal{V}_j^a} (P_{jk}^a \otimes \Phi_{jk}) \right. \\
\left. + \sum_{k \in \mathcal{V}_j^m} (P_{jk}^m \otimes \Phi_{jk}) \right] \mathfrak{N}(\mathbb{W}_j^t) \\
+ \sum_{k \in \mathcal{V}_j^d} (M_{jk} \otimes \Phi_{jk}) \mathfrak{N}(\mathbb{W}_k^t) = \mathbf{b}_j.
\end{aligned}$$

La matrice diagonale est ici formée de  $d_j^2$  blocs de taille  $6 \times 6$  et la matrice extradiagonale est formée de  $d_j \times d_k$  blocs de taille  $6 \times 6$ . Nous préférons utiliser la première formulation qui utilise un rangement par coordonnées puis par degré de liberté local. En posant  $\mathbf{W}_j = \mathfrak{N}(\mathbb{W}_j)$ , nous devons étudier le schéma :

$$\begin{aligned}
\left[ 2i\omega(\Phi_j \otimes \theta_j) + \sum_{l=1}^3 (\Phi_j^{x_l} \otimes G_{x_l}) + \sum_{k \in \mathcal{V}_j^a} (\Phi_{jk} \otimes P_{jk}^a) + \sum_{k \in \mathcal{V}_j^m} (\Phi_{jk} \otimes P_{jk}^m) \right] \mathbf{W}_j \\
+ \sum_{k \in \mathcal{V}_j^d} (\Phi_{jk} \otimes M_{jk}) \mathbf{W}_k = \mathbf{b}_j.
\end{aligned} \tag{103}$$

### 4.3 Étude du problème discret

#### 4.3.1 Formulation $P_1$

On se place maintenant dans le cas où  $\mathcal{P}_j = \mathbb{P}_1[C_j]$  et où les cellules  $C_j$  sont des tétraèdres (i.e.  $C_j = \tau_j$ ). Dans ce cas, on prend  $\phi_j = (\varphi_{j1}, \varphi_{j2}, \varphi_{j3}, \varphi_{j4}) = (\lambda_{j1}, \lambda_{j2}, \lambda_{j3}, \lambda_{j4})$ , où les  $\lambda_{jk}$  sont les coordonnées barycentriques relatives aux quatre sommets du tétraèdre  $\tau_j$ . En utilisant la formule :

$$\int_{\hat{\tau}} \lambda_1^{e_1} \lambda_2^{e_2} \lambda_3^{e_3} \lambda_4^{e_4} dx = \frac{e_1! e_2! e_3! e_4!}{(e_1 + e_2 + e_3 + e_4 + 3)!},$$

où  $\hat{\tau}$  désigne le tétraèdre de référence de volume  $\hat{V} = 1/6$ , on obtient :

$$\Phi_j = \int_{\tau_j} \phi_l^t \phi_m dx = \frac{V_j}{\hat{V}} \int_{\hat{\tau}} \hat{\phi}_l^t \hat{\phi}_m dx = \frac{V_j}{120\hat{V}} \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}, \tag{104}$$

où  $V_j$  désigne le volume du tétraèdre  $\tau_j$ . De même :

$$\Phi_j^{x_k} = \int_{\tau_j} \left( \phi_j^t (\partial_{x_k} \phi_j) - (\partial_{x_k} \phi_j)^t \phi_j \right) dx = \left( \frac{n_{mk} - n_{lk}}{12} \right)_{lm},$$

où  $n_{lk}$  est la composante selon  $e_k$  de la normale intérieure à  $\tau_j$  intégrée sur la face opposée au sommet numéro  $l$  dans  $\tau_j$ . Par ailleurs :

$$\Phi_{jk} = \int_{\Sigma_{jk}} \phi_j^t \phi_k ds$$

est une matrice  $4 \times 4$  dont le terme  $(lm)$  vaut  $\frac{\text{mes}(\Sigma_{jk})\alpha_{lm}}{12}$ , où  $\alpha_{lm}$  dépend des degrés de liberté associés (si le sommet  $l$  dans  $\tau_j$  n'est pas dans  $\tau_k$  ou réciproquement,  $\alpha_{lm} = 0$ ; s'ils sont égaux,  $\alpha_{lm} = 2$ ; sinon  $\alpha_{lm} = 1$ ). On a aussi :

$$\Phi_j \otimes \theta_j = \frac{V_j}{120\hat{V}} \begin{pmatrix} 2\theta_j & \theta_j & \theta_j & \theta_j \\ \theta_j & 2\theta_j & \theta_j & \theta_j \\ \theta_j & \theta_j & 2\theta_j & \theta_j \\ \theta_j & \theta_j & \theta_j & 2\theta_j \end{pmatrix} \quad \text{avec} \quad \theta_j = \begin{pmatrix} \varepsilon_j \mathbf{I}_3 & 0_3 \\ 0_3 & \mu_j \mathbf{I}_3 \end{pmatrix}. \quad (105)$$

La matrice  $\sum_{l=1}^3 \Phi_i^{x_l} \otimes G_{x_l}$  est explicitée dans l'annexe 7.7.3.

#### 4.3.2 Étude de la dispersion numérique

Tout comme pour le schéma VF (GDP0), nous allons procéder à l'étude de dispersion du schéma GDP1 sur une grille cartésienne (et non pas tétraédrique). La différence avec l'étude concernant le schéma GDP0 est que nous allons être contraints d'étudier un problème aux valeurs propres afin de trouver la relation de dispersion du schéma GDP1. Pour ce faire, nous considérerons les mêmes hypothèses que pour l'étude du schéma GDP0 à savoir : une grille régulière homogène dont les cellules sont de taille  $\Delta x \times \Delta y \times \Delta z$ , et une cellule  $C_{j,l,m}$  sur laquelle nous allons appliquer notre schéma et dont le centre de gravité  $G$  a pour coordonnées  $(j\Delta x, l\Delta y, m\Delta z)$ . Pour appliquer des fonctions  $\mathbb{P}_1$  sur un maillage hexaédrique orthogonal, nous considérons les fonctions de base suivantes :

$$\left\{ \begin{array}{llll} \varphi_1 = (1, 0, 0)^t, & \varphi_2 = (\frac{y-y_G}{\Delta y}, 0, 0)^t, & \varphi_3 = (\frac{z-z_G}{\Delta z}, 0, 0)^t, & \varphi_4 = (\frac{x-x_G}{\Delta x}, 0, 0)^t, \\ \varphi_5 = (0, 1, 0)^t, & \varphi_6 = (0, \frac{x-x_G}{\Delta x}, 0)^t, & \varphi_7 = (0, \frac{z-z_G}{\Delta z}, 0)^t, & \varphi_8 = (0, \frac{y-y_G}{\Delta y}, 0)^t, \\ \varphi_9 = (0, 0, 1)^t, & \varphi_{10} = (0, 0, \frac{x-x_G}{\Delta x})^t, & \varphi_{11} = (0, 0, \frac{y-y_G}{\Delta y})^t, & \varphi_{12} = (0, 0, \frac{z-z_G}{\Delta z})^t. \end{array} \right.$$

De ce fait, tout champ  $U_{j,l,m}$  désignant indifféremment  $E_{j,l,m}$  ou  $H_{j,l,m}$  sur une cellule  $C_{j,l,m}$  du maillage s'écrit :

$$\mathbf{U}_{j,l,m} = \begin{pmatrix} U_{j,l,m}^1 + U_{j,l,m}^4 \left( \frac{x - x_G}{\Delta x} \right) + U_{j,l,m}^2 \left( \frac{y - y_G}{\Delta y} \right) + U_{j,l,m}^3 \left( \frac{z - z_G}{\Delta z} \right) \\ U_{j,l,m}^5 + U_{j,l,m}^6 \left( \frac{x - x_G}{\Delta x} \right) + U_{j,l,m}^8 \left( \frac{y - y_G}{\Delta y} \right) + U_{j,l,m}^7 \left( \frac{z - z_G}{\Delta z} \right) \\ U_{j,l,m}^9 + U_{j,l,m}^{10} \left( \frac{x - x_G}{\Delta x} \right) + U_{j,l,m}^{11} \left( \frac{y - y_G}{\Delta y} \right) + U_{j,l,m}^{12} \left( \frac{z - z_G}{\Delta z} \right) \end{pmatrix}.$$

Nous partons donc de l'équation (95) en exploitant les hypothèses précédentes :

$$\begin{aligned} i\omega \int_{C_{i,j,k}} \boldsymbol{\varphi} \mathbf{W}_{j,l,m} dx &+ \frac{1}{2} \theta^{-1} \int_{C_{j,l,m}} \left( \boldsymbol{\varphi} \left( \sum_{r=1}^3 \partial_{x_r} G_{x_r} \mathbf{W}_{j,l,m} \right) - \sum_{r=1}^3 \partial_{x_r} \boldsymbol{\varphi} G_{x_r} \mathbf{W}_{j,l,m} \right) dx \\ &+ \frac{1}{2} \theta^{-1} \left( \int_{\Sigma_{i+1}} \boldsymbol{\varphi} M_{j+1,l,m} \mathbf{W}_{j+1,l,m} ds + \right. \\ &\quad \int_{\Sigma_{i-1}} \boldsymbol{\varphi} M_{j-1,l,m} \mathbf{W}_{j-1,l,m} ds + \\ &\quad \int_{\Sigma_{j+1}} \boldsymbol{\varphi} M_{j,l+1,m} \mathbf{W}_{j,l+1,m} ds + \\ &\quad \int_{\Sigma_{j-1}} \boldsymbol{\varphi} M_{j,l-1,m} \mathbf{W}_{j,l-1,m} ds + \\ &\quad \int_{\Sigma_{l+1}} \boldsymbol{\varphi} M_{j,l,m+1} \mathbf{W}_{j,l,m+1} ds + \\ &\quad \left. \int_{\Sigma_{l-1}} \boldsymbol{\varphi} M_{j,l,m-1} \mathbf{W}_{j,l,m-1} ds \right) = 0, \end{aligned}$$

qui donne les équations suivantes :



$$\begin{aligned}
i\omega \int_{C_{i,j,k}} \varphi \mathbf{E}_{j,l,m} dx &+ \frac{1}{2\mu} \int_{C_{j,l,m}} \left( \varphi \left( \sum_{r=1}^3 \partial_{x_r} N^{x_r} \mathbf{H}_{j,l,m} \right) - \sum_{r=1}^3 \partial_{x_r} \varphi N^{x_r} \mathbf{H}_{j,l,m} \right) dx \\
&+ \frac{1}{2\mu} \left( \int_{\Sigma_{i+1}} \varphi N_{j+1,l,m} \mathbf{H}_{j+1,l,m} ds + \right. \\
&\quad \int_{\Sigma_{i-1}} \varphi N_{j-1,l,m} \mathbf{H}_{j-1,l,m} ds + \\
&\quad \int_{\Sigma_{j+1}} \varphi N_{j,l+1,m} \mathbf{H}_{j,l+1,m} ds + \\
&\quad \int_{\Sigma_{j-1}} \varphi N_{j,l-1,m} \mathbf{H}_{j,l-1,m} ds + \\
&\quad \int_{\Sigma_{l+1}} \varphi N_{j,l,m+1} \mathbf{H}_{j,l,m+1} ds + \\
&\quad \left. \int_{\Sigma_{l-1}} \varphi N_{j,l,m-1} \mathbf{H}_{j,l,m-1} ds \right) = 0.
\end{aligned}$$

On remplace alors les champs du précédent système par leurs solutions particulières qui sont de la forme  $\mathbf{U}_{j,l,m} = \sum_{p=1}^{12} U_0^p e^{-i\mathbf{k} \cdot (i\Delta x, j\Delta y, l\Delta z)} \varphi_p$ . Si on note  $\tilde{\mathbf{U}} = (U_0^1, U_0^2, \dots, U_0^{12})^t$  on obtient le système d'équations :

$$\begin{cases} i\omega \tilde{\mathbf{E}} = \mu^{-1} \tilde{\mathbf{T}} \tilde{\mathbf{H}} \\ i\omega \tilde{\mathbf{H}} = -\varepsilon^{-1} \tilde{\mathbf{T}} \tilde{\mathbf{E}} \end{cases} \Rightarrow \omega^2 \tilde{\mathbf{E}} = c^2 \tilde{\mathbf{T}}^2 \tilde{\mathbf{E}}$$

avec :

$$\tilde{\mathbf{T}} = \begin{pmatrix} 0 & 0 & 0 & 0 & -s_z & 0 & -r_z & 0 & s_y & 0 & r_y & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -s_z & -12r_y & 0 & -3s_y & 0 \\ 0 & 0 & 0 & 0 & 12r_z & 0 & 3s_z & 0 & 0 & 0 & 0 & s_y \\ 0 & 0 & 0 & 0 & 0 & -s_z & 0 & 0 & 0 & s_y & 0 & 0 \\ s_z & 0 & r_z & 0 & 0 & 0 & 0 & 0 & -s_x & -r_x & 0 & 0 \\ 0 & 0 & 0 & s_z & 0 & 0 & 0 & 0 & 12r_x & 3s_x & 0 & 0 \\ -12r_z & 0 & -3s_z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -s_x \\ 0 & s_z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -s_x & 0 \\ -s_y & -r_y & 0 & 0 & s_x & r_x & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -s_y & -12r_x & -3s_x & 0 & 0 & 0 & 0 & 0 & 0 \\ 12r_y & 3s_y & 0 & 0 & 0 & 0 & 0 & s_x & 0 & 0 & 0 & 0 \\ 0 & 0 & -s_y & 0 & 0 & 0 & s_x & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

où on a posé, pour  $v = x, y, z$  :

$$\begin{cases} s_v &= \frac{i \sin(k_v \Delta v)}{\Delta v}, \\ r_v &= \frac{1 - \cos(k_v \Delta v)}{2 \Delta v}. \end{cases}$$

On est donc amené à trouver les valeurs propres  $\lambda$  de la matrice  $\tilde{T}$ . Ensuite, ces valeurs propres vérifient :

$$\omega^2 = c^2 \lambda^2, \quad (106)$$

ce qui donne une relation de dispersion. En fait, la numérotation des degrés de liberté utilisée est plutôt mal choisie. En prenant l'ordre suivant :

$$\begin{cases} \varphi_1 = (1, 0, 0)^t, & \varphi_2 = (0, 1, 0)^t, & \varphi_3 = (0, 0, 1)^t, \\ \varphi_4 = (\frac{x - x_G}{\Delta x}, 0, 0)^t, & \varphi_5 = (0, \frac{x - x_G}{\Delta x}, 0)^t, & \varphi_6 = (0, 0, \frac{x - x_G}{\Delta x})^t, \\ \varphi_7 = (\frac{y - y_G}{\Delta y}, 0, 0)^t, & \varphi_8 = (0, \frac{y - y_G}{\Delta y}, 0)^t, & \varphi_9 = (0, 0, \frac{y - y_G}{\Delta y})^t, \\ \varphi_{10} = (\frac{z - z_G}{\Delta z}, 0, 0)^t, & \varphi_{11} = (0, \frac{z - z_G}{\Delta z}, 0)^t, & \varphi_{12} = (0, 0, \frac{z - z_G}{\Delta z})^t, \end{cases}$$

on obtient une nouvelle matrice  $\tilde{T}$  ayant les mêmes valeurs propres, qui s'écrit :

$$\tilde{T} = \left( \begin{array}{ccc|ccc|ccc|ccc} 0 & -s_z & s_y & 0 & 0 & 0 & 0 & 0 & r_y & 0 & -r_z & 0 \\ s_z & 0 & -s_x & 0 & 0 & -r_x & 0 & 0 & 0 & r_z & 0 & 0 \\ -s_y & s_x & 0 & 0 & r_x & 0 & -r_y & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & -s_z & s_y & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 12r_x & s_z & 0 & 3s_x & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -12r_x & 0 & -s_y & -3s_x & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & -12r_y & 0 & 0 & 0 & 0 & -s_z & -3s_y & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & s_z & 0 & -s_x & 0 & 0 & 0 \\ 12r_y & 0 & 0 & 0 & 0 & 0 & 3s_y & s_x & 0 & 0 & 0 & 0 \\ \hline 0 & 12r_z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3s_z & s_y \\ -12r_z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3s_z & 0 & -s_x \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -s_y & s_x & 0 \end{array} \right),$$

On introduit un paramètre  $h$  qui nous permet de poser :

$$\begin{cases} \Delta x &= \alpha_x h, \\ \Delta y &= \alpha_y h, \\ \Delta z &= \alpha_z h, \end{cases}$$

et on suppose que les paramètres  $\alpha_x$ ,  $\alpha_y$  et  $\alpha_z$  restent bornés. Les valeurs propres de  $\tilde{T}$  sont recherchées sous la forme :

$$\lambda(\tilde{T}) = |\mathbf{k}|(1 + ah + bh^2) \text{ avec } (a, b) \in \mathbb{R}^2.$$

Le but est de trouver  $a$  et  $b$  pour connaître un développement limité de  $\lambda$  quand  $h \rightarrow 0$ , et donc une relation de dispersion asymptotique quand  $h \rightarrow 0$ . Tout d'abord, on effectue un développement limité des coefficients de la matrice  $\tilde{T}$  par rapport à  $h$  :

$$\begin{cases} s_v &= ik_v \left( 1 - \frac{k_v^2(\alpha_v h)^2}{6} + O(k_v^4 h^4) \right) \\ r_v &= \frac{k_v^2(\alpha_v h)}{4} + O(k_v^4 h^4) \end{cases}$$

Maintenant, nous recherchons les valeurs de  $a$  et de  $b$  pour lesquelles le polynôme caractéristique de  $\tilde{T}$  s'annule pour tout  $h$ .  $\lambda$  sera de la forme :

$$\lambda = |k| \left( 1 + \frac{1}{48} \frac{k_x^4 \alpha_x^2 + k_y^4 \alpha_y^2 + k_z^4 \alpha_z^2}{|k|^2} h^2 \right). \quad (107)$$

Ainsi, en injectant cette valeur propre dans l'équation (106), nous obtenons la relation de dispersion discrète :

$$\frac{\omega^2}{\omega_{ex}^2} = 1 + \frac{k_x^4 \Delta x^2 + k_y^4 \Delta y^2 + k_z^4 \Delta z^2}{24|k|^2} + O(|k|^4 h^4), \quad (108)$$

qui reste une dispersion d'ordre deux, huit fois moindre que celle qui caractérise le schéma volumes finis (86). Pour connaître par exemple la relation de dispersion en 1D, il suffit de prendre une onde qui se propage dans la direction  $\mathbf{k} = |\mathbf{k}|(1, 0, 0)$ , ce qui donne :

$$\frac{\omega^2}{\omega_{ex}^2} = 1 + \frac{|k|^2 \Delta x^2}{24} + O(|k|^4 h^4). \quad (109)$$

## 5 Résultats numériques

### 5.1 Résultats 1D

On considère le domaine  $[a, b] = [0, 1]$ . Conformément aux conditions que nous avons choisies dans les paragraphes précédents, nous optons pour une condition absorbante aux points  $\{a\}$  et  $\{b\}$ . Dans le cas test considéré, une onde plane se dirige du point  $\{a\}$  vers le point  $\{b\}$ , et on observe la réponse obtenue pour différentes caractéristiques du maillage et du milieu. L'onde plane est de la forme  $W^{\text{inc}}(x) = (\eta_e, \eta_h)e^{-ikx}$  où  $k = \frac{\omega}{c_1}$  et  $\eta_h = z_1 \eta_e^4$ , de sorte qu'elle satisfait bien les équations de Maxwell 1D. Le maillage est toujours construit de façon à ce que l'on ait un minimum de 10 points par longueur d'onde. Pour chaque test, on compare les solutions fournies par les versions GDP0 et GDP1. On choisit de visualiser la partie réelle du champ électrique.

<sup>4</sup> $c_1$  et  $z_1$  sont les valeurs de  $c$  et  $z$  dans la cellule  $C_1$ .

### 5.1.1 Maillage uniforme, milieu homogène

Dans ce test, on pose pour  $\forall j : \varepsilon_j = 1$  et  $\mu_j = 1$ . Les conditions en  $\{a\}$  et  $\{b\}$  permettent de simuler la propagation d'une onde plane dans le vide et donc de pouvoir comparer la solution obtenue à la solution de référence. Pour l'onde plane choisie, cette solution est  $\Re(e^{-ikx}) = \cos(kx)$ . On utilise un maillage uniforme de  $N = 200$  cellules pour les fréquences  $f = 100$  MHz,  $f = 500$  MHz et  $f = 1$  GHz. Rappelons au passage les relations de base :

$$\begin{cases} k &= \frac{\omega}{c}, \\ \omega &= 2\pi f, \\ \lambda &= \frac{c}{f}. \end{cases}$$

dont on tire  $\lambda = \frac{2\pi}{k}$ . Les solutions obtenues sont montrées sur la figure 3. En comparant ces graphiques, on constate que la solution P1 est plus précise que la solution P0, en particulier pour les hautes fréquences.

On souhaite maintenant évaluer l'impact de la dispersion numérique, toujours en maillage uniforme et en milieu homogène. Ces conditions permettent de considérer la relation de dispersion trouvée pour la méthode GDP1, qui s'exprime à partir de la relation 3D (109) qui conduit à :

$$|\omega^2 - \omega_h^2| \sim \omega_h^2 \frac{k_x^2 \Delta x^2}{24},$$

qui montre que plus le nombre de points par longueur d'onde ( $\frac{\lambda}{\Delta x} = \frac{2\pi}{k_x \Delta x}$ ) sera petit, plus l'erreur de dispersion sera importante. Pour garantir une précision acceptable, il faut donc s'assurer que la taille de maille  $\Delta x$  corresponde à un certain nombre de points par longueur d'onde. Dans le cas présent, nous avons fixé ce nombre à 10. Pour noter cette différence, on peut par exemple fixer une taille de maille et augmenter la fréquence le plus possible tout en satisfaisant le critère de 10 points par longueur d'onde. Ainsi, en prenant une fréquence  $f = 1$  GHz et un maillage de  $N = 50$  cellules, on obtient la courbe de la figure 4.

La solution résultant de la méthode GDP1 permet de noter clairement l'erreur de dispersion pour ce jeu de données, alors que la solution associée à la méthode GDP0 est trop altérée pour s'en rendre compte. Ici, cette erreur est caractérisée comme prévu par un déphasage entre les deux courbes. Si on souhaite comparer les deux méthodes, il faut considérer le cas d'un maillage de  $N = 500$  cellules et faire un zoom (cf. figure 5). On constate une erreur de dispersion pour les deux méthodes bien que la différence ne soit pas vraiment significative. Pourtant, nous avons zoomé sur une zone où la solution associée à la méthode GDP0 était la plus proche de la solution exacte. Si on se réfère aux formules qui expriment la relation de dispersion pour la méthode GDP0 et la relation de dispersion pour la méthode GDP1, nous constatons que cette dernière est moins dispersive d'un facteur 8. Mais pour bien se rendre compte de cette erreur, il faudrait comparer les courbes en prenant un maillage très

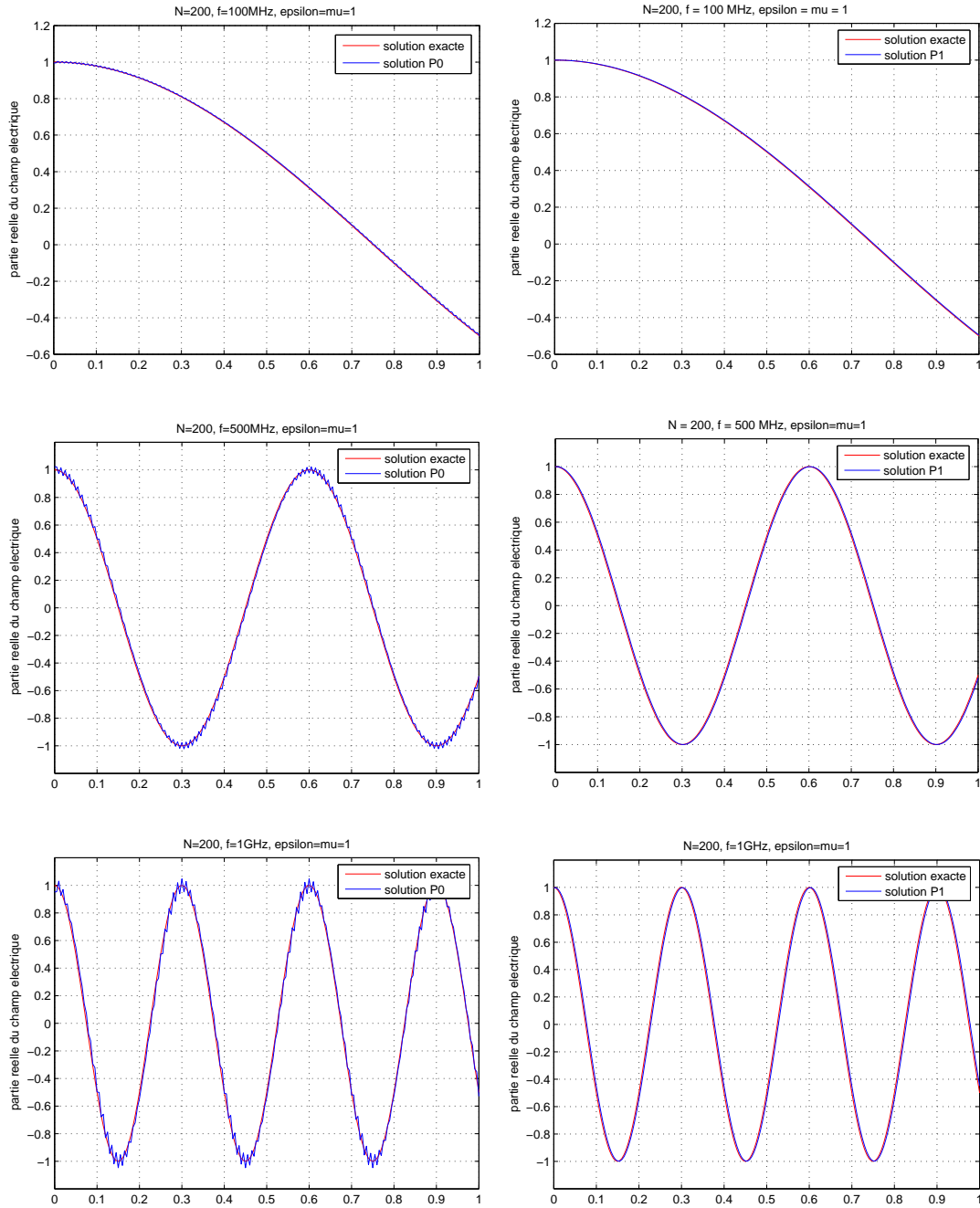


FIG. 3 – Equations de Maxwell 1D : milieu homogène et maillage uniforme

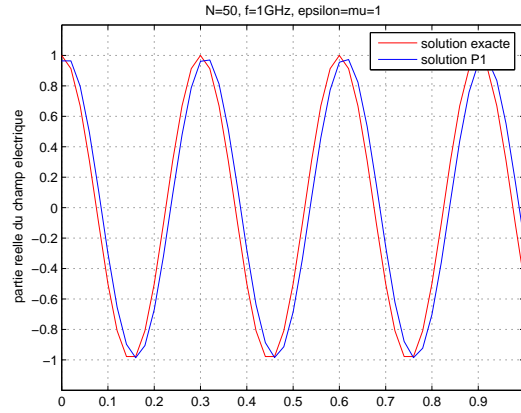


FIG. 4 – Equations de Maxwell 1D : milieu homogène et maillage uniforme  
Erreur de dispersion pour un maillage de  $N = 50$  cellules et  $f = 1$  GHz

grossier, chose qui est irréalisable compte tenu de la mauvaise qualité de la solution obtenue avec la méthode GDP0.

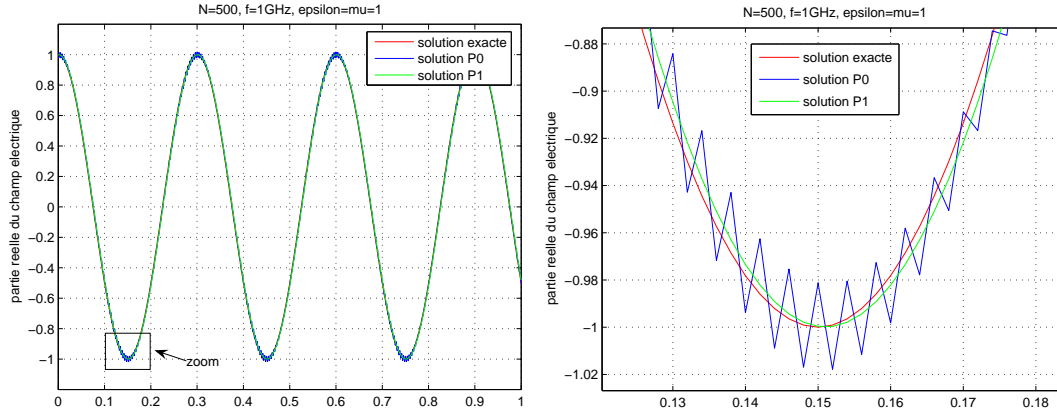


FIG. 5 – Equations de Maxwell 1D : milieu homogène et maillage uniforme  
Erreur de dispersion pour un maillage de  $N = 500$  cellules et  $f = 1$  GHz

### 5.1.2 Maillage non uniforme, milieu homogène

Dans cette section, le but est de voir comment se comporte la méthode GDP1 en présence d'un maillage non uniforme. Les conditions aux limites sont identiques à celles prises dans

la section 5.1.1. Le maillage comporte 200 cellules mais le pas d'espace est défini de façon aléatoire tout en vérifiant toujours le critère  $\max_{1 \leq j \leq N} (\Delta x_j) \leq \lambda/10$ . Les cellules sont de trois tailles différentes bien que la longueur de la plus grande maille reste égale à trois fois celle de la plus petite. Des tests sont réalisés pour des fréquences de 100 MHz, 500 MHz et 1 GHz. Les solutions obtenues sont montrées sur la figure 6. Là où la méthode GDP0 pêche par manque de précision, la solution résultant de la méthode GDP1 est correcte même sur un maillage aléatoire grossier. On peut donc prendre un maillage aléatoire de  $N = 200$  cellules et observer que la solution se calque sur la solution de référence.

Ceci dit, un test intéressant consiste à comparer les solutions P0 et P1 obtenues en considérant un maillage aléatoire fin pour la méthode GDP0, et un maillage aléatoire grossier pour la méthode GDP1. Le but est bien entendu d'évaluer la qualité de la solution obtenue, et de voir s'il est plus avantageux de raffiner le maillage ou d'augmenter l'ordre d'interpolation (cf. figure 7). Il est assez clair qu'il est beaucoup plus avantageux d'augmenter l'ordre d'interpolation de la méthode GD plutôt que de raffiner le maillage lorsqu'on recherche une solution précise. En dépit d'un maillage 6 fois plus fin dans les tests utilisant la méthode P0, la solution reste non seulement moins précise, mais elle est également plus dispersive. Néanmoins, ces remarques ne se soucient point de l'aspect résolution : l'opérateur matriciel issu de la méthode GDP1 est peut être moins bien conditionné que celui issu de la méthode GDP0. Cet aspect n'est pas abordé ici.

### 5.1.3 Maillage uniforme, milieu hétérogène

Dans cette section, toutes les conditions des essais numériques des sections 5.1.1 et 5.1.2 sont reproduites, sauf deux. D'une part, sur la section  $[0.4; 0.6]$  du domaine de calcul, on pose  $\varepsilon_j = 4$  et  $\mu_j = 1$ . Pour le reste du domaine, nous conserverons  $\varepsilon = \mu = 1$ . D'autre part, nous avons fixé un nombre de cellules à  $N = 400$  de sorte que la solution fournie par la méthode GDP0 ne subisse pas les inconvénients d'un maillage grossier. Ces tests sont réalisés pour des fréquences de 100 MHz, 500 MHz et 1 GHz. Les solutions obtenues sont montrées sur la figure 8. On observe que les solutions associées aux méthodes GDP0 et GDP1 ne coïncident absolument pas. Cette différence est d'autant plus notable que la fréquence est élevée.

### 5.1.4 Maillage non uniforme, milieu hétérogène

Les conditions de simulation sont strictement les mêmes que pour la section 5.1.3 hormis le fait que le maillage est non uniforme. Le nombre de cellules est  $N = 400$ . Les solutions obtenues sont montrées sur la figure 9. Nous constatons que les solutions associées aux méthodes GDP0 et GDP1 ont un comportement et un profil identique à leur homologue en maillage uniforme. On retrouve les oscillations de la solution P0 en milieu homogène, mais la solution P1 est quasi-identique à la solution P1 en maillage uniforme. Ces résultats confirment que la méthode GDP1 est bien moins sensible aux irrégularités de maillage et permet visiblement de mieux prendre en compte l'hétérogénéité. On remarque notamment que pour une fréquence  $f = 100$  MHz, le profil de la solution P1 change lorsqu'elle se situe

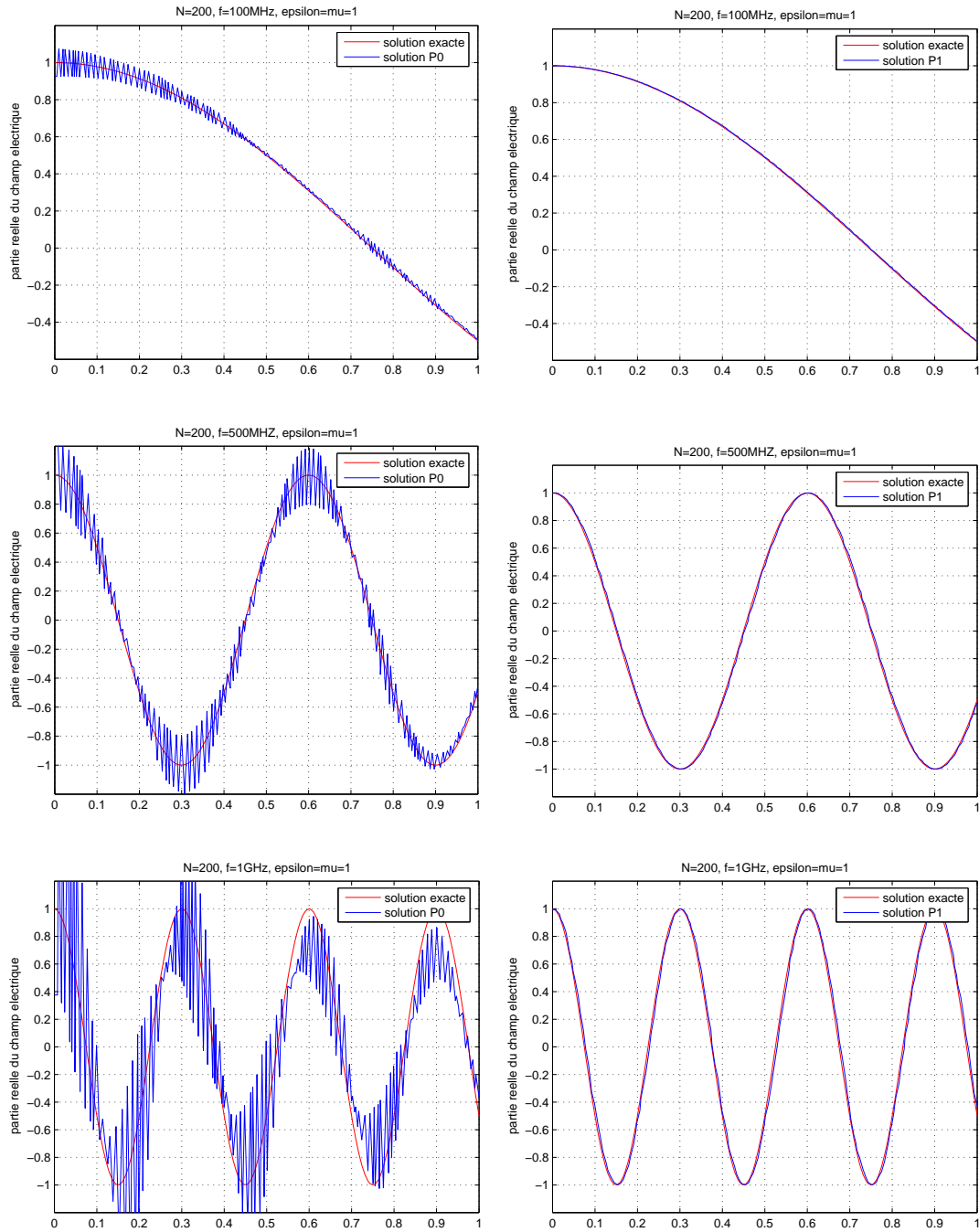


FIG. 6 – Equations de Maxwell 1D : milieu homogène et maillage non uniforme



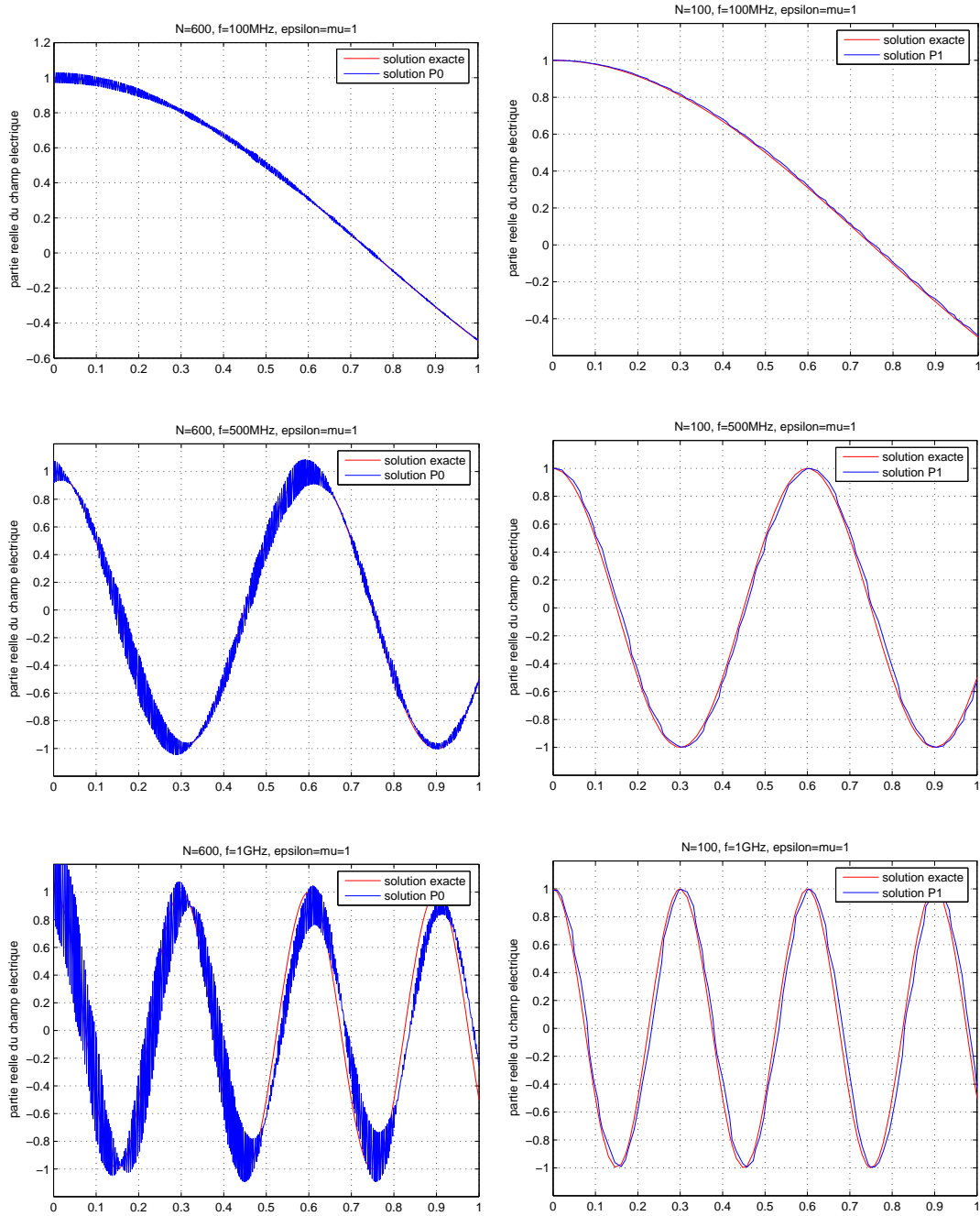


FIG. 7 – Equations de Maxwell 1D : milieu homogène et maillage non uniforme

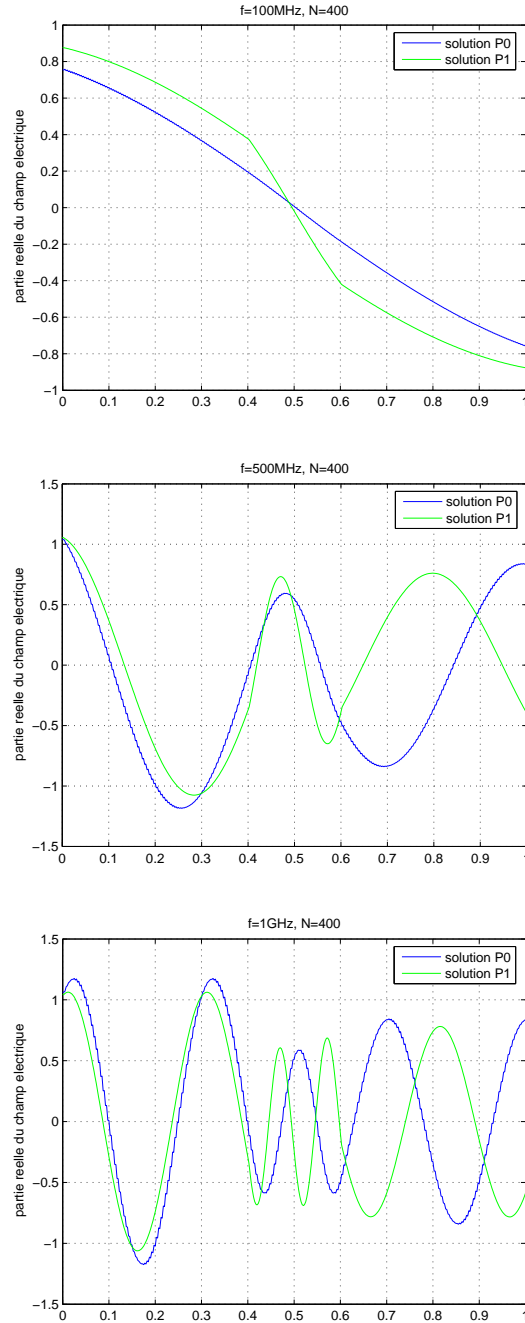


FIG. 8 – Equations de Maxwell 1D : milieu hétérogène et maillage uniforme

dans la zone hétérogène  $[0.4; 0.6]$ , alors que la solution P0 ne semble pas prendre en compte cette transition.

## 5.2 Résultats 3D

On présente ici des résultats préliminaires concernant la résolution numérique des équations de Maxwell 3D en maillage tétraédrique. Le cas test considéré est la diffraction d'une onde plane sur une sphère unitaire métallique. La fréquence de l'onde incidente est fixée à 300 MHz. Le maillage tétraédrique contient 43 932 sommets et 244 608 tétraèdres. La longueur des arêtes du maillage varie entre  $L_{\min}=0.054$  et  $L_{\max}=0.216$  (la taille moyenne est  $L_{\text{moy}}=0.108$ ). Le maillage utilisé est donc relativement uniforme. La limite artificielle du domaine de calcul, sur laquelle est imposée une condition absorbante, est fixée à une longueur d'onde de la surface de la sphère (cette frontière absorbante est une sphère de rayon  $R=2$ ). Le calcul est effectué en champ total. La méthode GDP0 conduit à un nombre de degrés de liberté égal à  $6 \times 244,608 = 1\,467\,648$  alors que pour la méthode GDP1 ce nombre est  $24 \times 244,608 = 5\,870\,592$ . Les systèmes linéaires obtenus sont résolus par une méthode BiCGstab( $\ell$ ) [36] non préconditionnée. La convergence de cette méthode est montrée sur la figure 10. Il est bien évident que la prise en compte d'une technique de préconditionnement adaptée aux propriétés des matrices (non hermitiennes) résultant des formulations GDP0 et GDP1 considérées ici s'impose pour une comparaison raisonnable de ces méthodes en termes d'efficacité. Les solutions obtenues sont montrées sur les figures 11 à 13 sous la formes des lignes de contour de la partie réelle des composantes du champ électrique. Ces solutions montrent une nette supériorité de la méthode GDP1. Néanmoins, il serait intéressant d'affiner cette comparaison en considérant différentes tailles de maillages et en confrontant les solutions obtenues avec une solution fournie par une méthode d'éléments finis frontière. Ceci est en cours de réalisation.

## 6 Conclusions et perspectives

Nous avons ici formulé et étudié les principales propriétés de deux méthodes de type Galerkin Discontinu centrées pour la résolution numérique des équations de Maxwell en domaine fréquentiel. Nous avons plus précisément considéré une méthode GD d'ordre 0 (GDP0 ou méthode de volumes finis) et une méthode GD d'ordre 1 (GDP1) basée sur une interpolation linéaire au sein de chaque élément du maillage. La mise en œuvre dans le cas 3D est réalisée sur des maillages tétraédriques arbitrairement non-structurés. Des résultats numériques préliminaires ont permis de valider les méthodes GDP0 et GDP1. Ces résultats ont clairement démontré l'intérêt de la formulation GDP1 lorsque le maillage est non uniforme. Il est fort probable que des formulations d'ordre élevé (GDP $k$ ,  $k \geq 2$ ) s'avèreront encore plus avantageuses. Nous n'avons pas traité la question de la résolution efficace des systèmes linéaires résultant qui doit être abordée en priorité avant d'envisager le passage à de telles formulations.

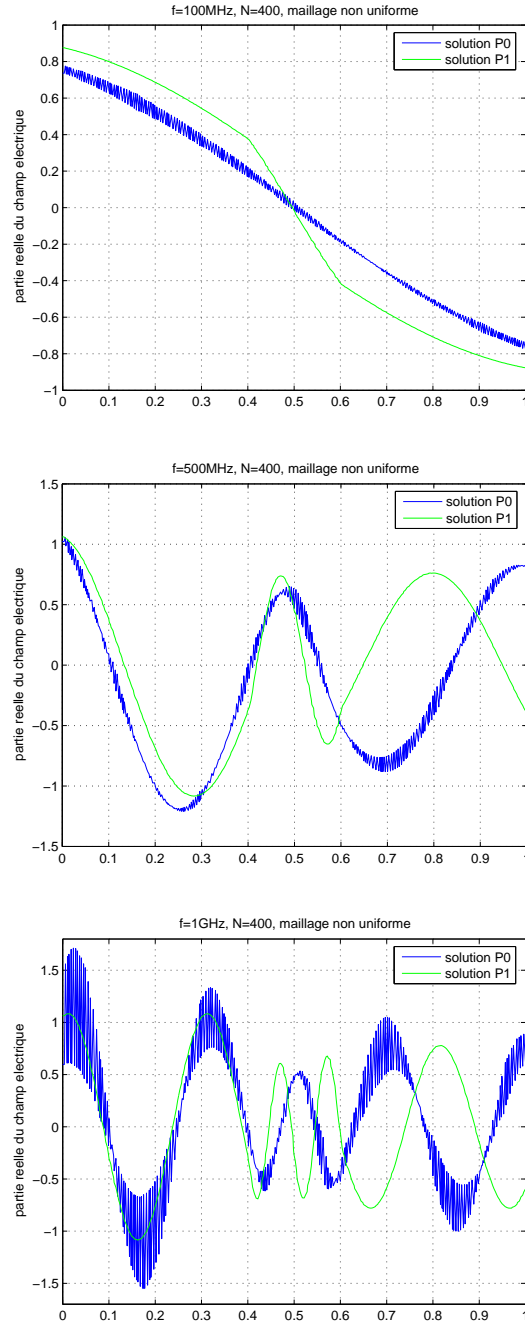


FIG. 9 – Equations de Maxwell 1D : milieu hétérogène et maillage non uniforme

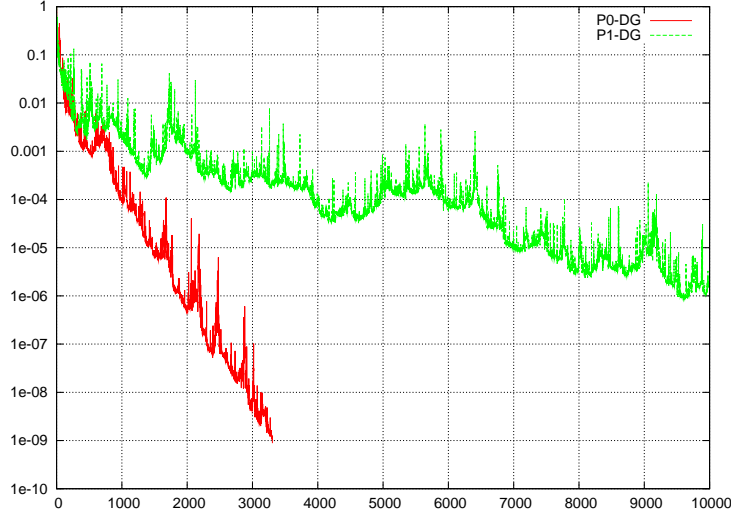


FIG. 10 – Equations de Maxwell 3D : diffraction par une sphère métallique  
Convergence de la méthode BiCGstab( $\ell$ )

**Remerciements.** Nos remerciements vont à Ronan Perrussel pour sa relecture minutieuse et ses commentaires constructifs sur une version préliminaire de ce rapport.

## 7 Annexes

### 7.1 Équivalences des conditions aux limites 1D

**Proposition 4** *Pour imposer faiblement la condition au bord absorbant 1D, les deux égalités suivantes sont équivalentes :*

$$(i) E_{\frac{1}{2}} - z_1 H_{\frac{1}{2}} = \xi(E^{inc} - z_1 H^{inc}),$$

$$(ii) A_1 \frac{W_0 + W_1}{2} = \xi A_1 W^{inc},$$

$$\text{où } A_1 = \begin{pmatrix} 1 & -z_1 \\ -z_1^{-1} & 1 \end{pmatrix}.$$

**Preuve** On part dans un premier temps de (i) que l'on multiplie par  $-z_1^{-1}$  de manière à obtenir :

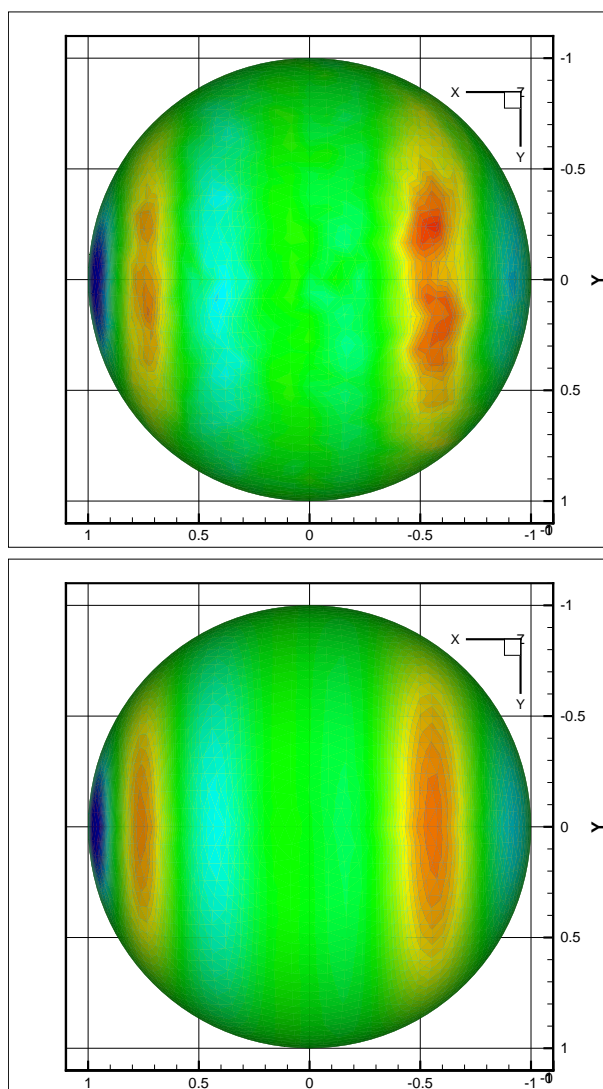


FIG. 11 – Equations de Maxwell 3D : diffraction par une sphère métallique  
Contours de  $\Re(E_x)$  : méthode GDP0 (haut) et GDP1 (bas)

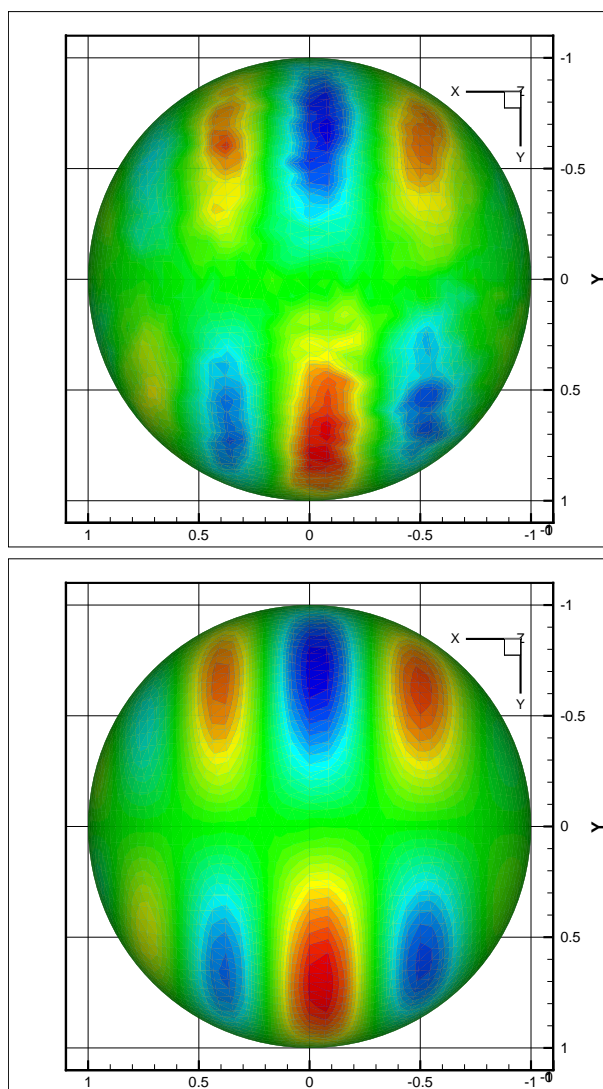


FIG. 12 – Equations de Maxwell 3D : diffraction par une sphère métallique  
Contours de  $\Re(E_y)$  : méthode GDP0 (haut) et GDP1 (bas)

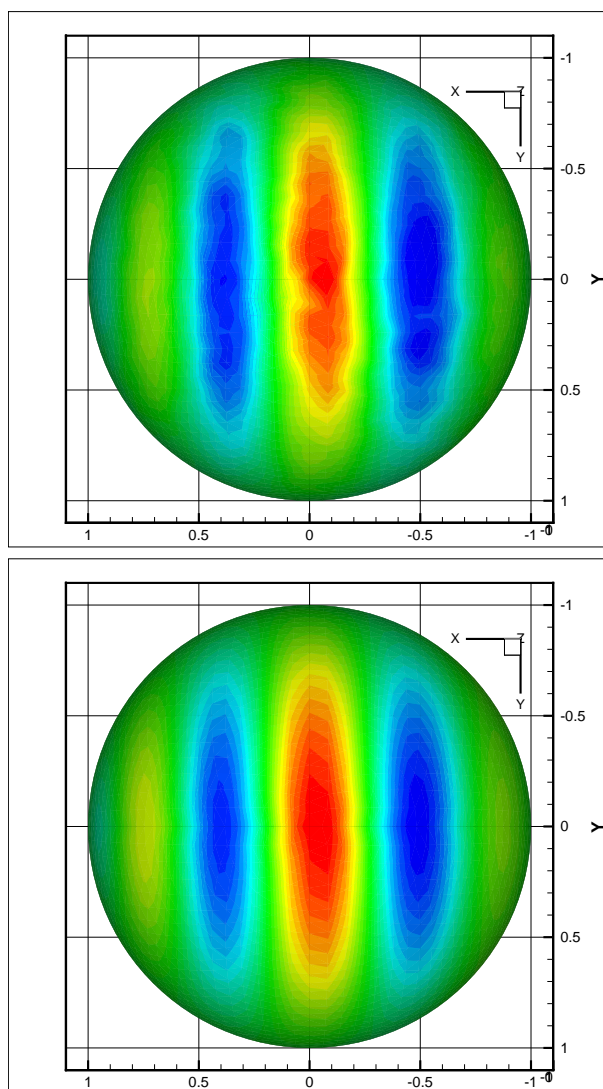


FIG. 13 – Equations de Maxwell 3D : diffraction par une sphère métallique  
Contours de  $\Re(E_y)$  : méthode GDP0 (haut) et GDP1 (bas)



$$\begin{cases} E_{\frac{1}{2}} - z_1 H_{\frac{1}{2}} &= \xi(E^{\text{inc}} - z_1 H^{\text{inc}}) \\ -z_1^{-1} E_{\frac{1}{2}} + H_{\frac{1}{2}} &= \xi(-z_1^{-1} E^{\text{inc}} + H^{\text{inc}}) \end{cases}$$

$$\Leftrightarrow \begin{pmatrix} 1 & -z_1 \\ -z_1^{-1} & 1 \end{pmatrix} \begin{pmatrix} E_{\frac{1}{2}} \\ H_{\frac{1}{2}} \end{pmatrix} = \xi \begin{pmatrix} 1 & -z_1 \\ -z_1^{-1} & 1 \end{pmatrix} \begin{pmatrix} E^{\text{inc}} \\ H^{\text{inc}} \end{pmatrix},$$

ce qui correspond à :

$$A_1 \frac{W_0 + W_1}{2} = \xi A_1 W^{\text{inc}}.$$

■

**Proposition 5** *Pour imposer faiblement la condition au bord métallique 1D, les deux égalités suivantes sont équivalentes :*

$$(i) E_{N+\frac{1}{2}} = (\xi - 1)E^{\text{inc}},$$

$$(ii) A_N \frac{W_N + W_{N+1}}{2} = (\xi - 1)A_N W^{\text{inc}},$$

$$\text{où } A_N = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

**Preuve** Si on utilise le fait que  $E_{N+\frac{1}{2}} = \frac{E_N + E_{N+1}}{2}$ , on a en partant de (i) :

$$E_{N+\frac{1}{2}} = (\xi - 1)E^{\text{inc}} \Leftrightarrow \frac{E_N + E_{N+1}}{2} = (\xi - 1)E^{\text{inc}}$$

$$\Leftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{E_N + E_{N+1}}{2} \\ \frac{H_N + H_{N+1}}{2} \end{pmatrix} = (\xi - 1) \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} E^{\text{inc}} \\ H^{\text{inc}} \end{pmatrix},$$

ce qui exprime exactement :

$$A_N \frac{W_N + W_{N+1}}{2} = (\xi - 1)A_N W^{\text{inc}}.$$

■

## 7.2 Propriétés de la matrice $N_{jk}$

**Proposition 6** *La matrice  $N_{jk}$  possède les propriétés suivantes :*

(i)  $N_{jk}$  est anti-symétrique et  $N_{jk} = -N_{kj}$ .

$$(ii) \sum_{k \in \mathcal{V}_j} S_{jk} N_{jk} = 0.$$

(iii) Soit  $\mathbf{X} \in \mathbb{C}^3$ . Si on note respectivement  $\mathbf{X}_\perp$  et  $\mathbf{X}_\parallel$  les parties normales et tangentielles de  $\mathbf{X}$  relativement à la surface  $\Sigma_{jk}$  :

$$\mathbf{X}_\perp = (\mathbf{n}_{jk} \mathbf{n}_{jk}^t) \mathbf{X} = (\mathbf{n}_{jk} \cdot \mathbf{X}) \mathbf{n}_{jk} \quad \text{et} \quad \mathbf{X}_\parallel = \mathbf{n}_{jk} \times (\mathbf{X} \times \mathbf{n}_{jk}) = \mathbf{X} - (\mathbf{n}_{jk} \cdot \mathbf{X}) \mathbf{n}_{jk},$$

$$\text{alors on a que } \mathbf{X}_\parallel = -N_{jk}^2 \mathbf{X} \text{ et donc que } \mathbf{X}_\perp = (N_{jk}^2 + I_3) \mathbf{X}.$$

$$(iv) N_{jk}^3 = -N_{jk}.$$

(v)  $N_{jk}^2$  est symétrique et négative et elle admet les valeurs propres  $\{0, -1, -1\}$ .

#### Preuve

(i) Tout d'abord,  $N_{ji} \mathbf{X} = \mathbf{X} \times \mathbf{n}_{ji} = -\mathbf{X} \times \mathbf{n}_{jk} = -N_{jk} \mathbf{X} \Rightarrow N_{ji} = -N_{jk}$ .

(ii) En faisant les calculs on obtient :

$$\begin{aligned} \sum_{j \in \mathcal{V}_j} S_{jk} \mathbf{n}_{jk} = 0 & \Rightarrow \sum_{j \in \mathcal{V}_j} S_{jk} N_{jk} \mathbf{X} = - \left( \sum_{j \in \mathcal{V}_j} S_{jk} \mathbf{n}_{jk} \right) \times \mathbf{X} = 0 \\ & \Rightarrow \sum_{j \in \mathcal{V}_j} S_{jk} N_{jk} = 0. \end{aligned}$$

(iv) En faisant les calculs on obtient :

$$\begin{aligned} N_{jk}^3 \mathbf{X} &= ((\mathbf{X} \times \mathbf{n}_{jk}) \times \mathbf{n}_{jk}) \times \mathbf{n}_{jk} \\ &= -(\mathbf{n}_{jk} \times (\mathbf{X} \times \mathbf{n}_{jk})) \times \mathbf{n}_{jk} \\ &= -(\mathbf{X} - (\mathbf{n}_{jk} \cdot \mathbf{X}) \mathbf{n}_{jk}) \times \mathbf{n}_{jk} \\ &= -\mathbf{X} \times \mathbf{n}_{jk} \\ &= -N_{jk} \mathbf{X} \\ \Rightarrow N_{jk}^3 &= -N_{jk}. \end{aligned}$$

■

### 7.3 Décomposition de la matrice des flux

On cherche dans ce paragraphe à décomposer la matrice  $M_{jk}^j$  en sa partie positive et négative. Nous avons déjà vu que :

$$M_{jk}^j = M_{jk} \theta_j^{-1} \text{ avec } M_{jk} = \begin{pmatrix} 0_3 & N_{jk} \\ -N_{jk} & 0_3 \end{pmatrix} \text{ et } \theta_j = \begin{pmatrix} \varepsilon_j \mathbf{I}_3 & 0_3 \\ 0_3 & \mu_j \mathbf{I}_3 \end{pmatrix}.$$

L'hyperbolicité du système d'équations de Maxwell signifie que  $M_{jk}^j$  est diagonalisable. On montre que  $M_{jk}^j$  admet les valeurs propres réelles doubles  $\{\lambda_0 = 0, \lambda_+ = c_j, \lambda_- = -c_j\}$ . Nous allons d'abord trouver une paire de vecteurs propres orthonormaux associés à la valeur propre  $\lambda_0$ . Si on choisit la paire de vecteurs  $(\mathbf{W}_0^a, \mathbf{W}_0^b) \in \mathbb{R}^{6 \times 2}$  telle que  $\mathbf{W}_0^a = (\mathbf{n}_{jk}, 0_3)^t$  et  $\mathbf{W}_0^b = (0_3, \mathbf{n}_{jk})^t$ , on s'aperçoit qu'ils sont orthonormaux entre eux, et qu'ils correspondent bien à des vecteurs propres de  $M_{jk}^j$  associés à  $\lambda_0$ . En effet :

$$\begin{cases} M_{jk}^j \mathbf{W}_0^a = M_{jk} \theta_j^{-1} \mathbf{W}_0^a = \begin{pmatrix} 0_3 & N_{jk} \\ -N_{jk} & 0_3 \end{pmatrix} \begin{pmatrix} \varepsilon_j^{-1} \mathbf{n}_{jk} \\ 0_3 \end{pmatrix} = \begin{pmatrix} 0_3 \\ -\varepsilon_j^{-1} N_{jk} \mathbf{n}_{jk} \end{pmatrix} = 0_6 \\ M_{jk}^j \mathbf{W}_0^b = M_{jk} \theta_j^{-1} \mathbf{W}_0^b = \begin{pmatrix} 0_3 & N_{jk} \\ -N_{jk} & 0_3 \end{pmatrix} \begin{pmatrix} 0_3 \\ \mu_j^{-1} \mathbf{n}_{jk} \end{pmatrix} = \begin{pmatrix} -\mu_j^{-1} N_{jk} \mathbf{n}_{jk} \\ 0_3 \end{pmatrix} = 0_6. \end{cases}$$

On peut par conséquent définir le projecteur orthogonal  $\Pi_{jk}^0$  sur le sous-espace propre  $\ker(M_{jk}^j)$  par :

$$\Pi_{jk}^0 = \mathbf{W}_0^a (\mathbf{W}_0^a)^t + \mathbf{W}_0^b (\mathbf{W}_0^b)^t = \begin{pmatrix} \mathbf{n}_{jk} \mathbf{n}_{jk}^t & 0_3 \\ 0_3 & \mathbf{n}_{jk} \mathbf{n}_{jk}^t \end{pmatrix} = \begin{pmatrix} N_{jk}^2 + \mathbf{I}_3 & 0_3 \\ 0_3 & N_{jk}^2 + \mathbf{I}_3 \end{pmatrix}.$$

On note  $\Pi_{jk}^+$  et  $\Pi_{jk}^-$  les projecteurs orthogonaux sur les sous-espaces propres respectifs  $\ker((M_{jk}^j - c_j \mathbf{I}_6))$  et  $\ker((M_{jk}^j + c_j \mathbf{I}_6))$ . En raison de la diagonalisabilité de  $M_{jk}^j$ , on a :

$$\begin{cases} \mathbf{I}_6 &= \Pi_{jk}^0 + \Pi_{jk}^+ + \Pi_{jk}^- \\ M_{jk}^j &= \lambda_0 \Pi_{jk}^0 + \lambda_+ \Pi_{jk}^+ + \lambda_- \Pi_{jk}^- \end{cases} \Leftrightarrow \begin{cases} \mathbf{I}_6 &= \Pi_{jk}^0 + \Pi_{jk}^+ + \Pi_{jk}^- \\ M_{jk}^j &= c_j \Pi_{jk}^+ - c_j \Pi_{jk}^- \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} \Pi_{jk}^+ &= \frac{1}{2}(\mathbf{I}_6 - \Pi_{jk}^0 + c_j^{-1} M_{jk}^j) \\ \Pi_{jk}^- &= \frac{1}{2}(\mathbf{I}_6 - \Pi_{jk}^0 - c_j^{-1} M_{jk}^j). \end{cases}$$

On se sert maintenant de la matrice d'impédance  $Z_j$  si bien que l'on a :

$$c_j^{-1} \theta_j^{-1} = \begin{pmatrix} (c_j \varepsilon_j)^{-1} \mathbf{I}_3 & 0_3 \\ 0_3 & (c_j \mu_j)^{-1} \mathbf{I}_3 \end{pmatrix} = Z_j^{-1},$$

et de ce fait :

$$\begin{aligned}
\Pi_{jk}^+ &= \frac{1}{2}(\mathbf{I}_6 - \Pi_{jk}^0 + c_j^{-1} M_{jk} \theta_j^{-1}) \\
&= \frac{1}{2} \left[ \begin{pmatrix} \mathbf{I}_3 & 0_3 \\ 0_3 & \mathbf{I}_3 \end{pmatrix} - \begin{pmatrix} N_{jk}^2 + \mathbf{I}_3 & 0_3 \\ 0_3 & N_{jk}^2 + \mathbf{I}_3 \end{pmatrix} + M_{jk} Z_j^{-1} \right] \\
&= \frac{1}{2} \left[ \begin{pmatrix} -N_{jk}^2 & 0_3 \\ 0_3 & -N_{jk}^2 \end{pmatrix} + \begin{pmatrix} 0_3 & N_{jk} \\ -N_{jk} & 0_3 \end{pmatrix} \begin{pmatrix} z_j \mathbf{I}_3 & 0_3 \\ 0_3 & z_j^{-1} \mathbf{I}_3 \end{pmatrix} \right] \\
&= \frac{1}{2} \begin{pmatrix} -N_{jk}^2 & z_j^{-1} N_{jk} \\ -z_j N_{jk} & -N_{jk}^2 \end{pmatrix},
\end{aligned}$$

et :

$$\begin{aligned}
\Pi_{jk}^- &= \frac{1}{2}(\mathbf{I}_6 - \Pi_{jk}^0 - c_j^{-1} M_{jk} \theta_j^{-1}) \\
&= \frac{1}{2} \left[ \begin{pmatrix} \mathbf{I}_3 & 0_3 \\ 0_3 & \mathbf{I}_3 \end{pmatrix} - \begin{pmatrix} N_{jk}^2 + \mathbf{I}_3 & 0_3 \\ 0_3 & N_{jk}^2 + \mathbf{I}_3 \end{pmatrix} - M_{jk} Z_j^{-1} \right] \\
&= \frac{1}{2} \left[ \begin{pmatrix} -N_{jk}^2 & 0_3 \\ 0_3 & -N_{jk}^2 \end{pmatrix} + \begin{pmatrix} 0_3 & -N_{jk} \\ N_{jk} & 0_3 \end{pmatrix} \begin{pmatrix} z_j \mathbf{I}_3 & 0_3 \\ 0_3 & z_j^{-1} \mathbf{I}_3 \end{pmatrix} \right] \\
&= \frac{1}{2} \begin{pmatrix} -N_{jk}^2 & -z_j^{-1} N_{jk} \\ z_j N_{jk} & -N_{jk}^2 \end{pmatrix},
\end{aligned}$$

et par suite :  $\Pi_{jk}^+ = \frac{1}{2} \begin{pmatrix} -N_{jk}^2 & z_j^{-1} N_{jk} \\ -z_j N_{jk} & -N_{jk}^2 \end{pmatrix}$  et  $\Pi_{jk}^- = \frac{1}{2} \begin{pmatrix} -N_{jk}^2 & -z_j^{-1} N_{jk} \\ z_j N_{jk} & -N_{jk}^2 \end{pmatrix}$ .

Maintenant, on utilise l'expression de ces projecteurs pour en déduire les matrices  $(M_{jk}^j)^\pm$  avec, par définition  $M_{jk}^j = (M_{jk}^j)^+ + (M_{jk}^j)^-$  et vu que  $M_{jk}^j = c_j \Pi_{jk}^+ - c_j \Pi_{jk}^-$ , alors  $(M_{jk}^j)^+ = c_j \Pi_{jk}^+$  et  $(M_{jk}^j)^- = -c_j \Pi_{jk}^-$ . Donc :

$$(M_{jk}^j)^+ = \frac{c_j}{2} \begin{pmatrix} -N_{jk}^2 & z_j^{-1} N_{jk} \\ -z_j N_{jk} & -N_{jk}^2 \end{pmatrix} \quad \text{et} \quad (M_{jk}^j)^- = \frac{c_j}{2} \begin{pmatrix} N_{jk}^2 & z_j^{-1} N_{jk} \\ -z_j N_{jk} & N_{jk}^2 \end{pmatrix}.$$

On notera par ailleurs que  $(M_{jk}^j)^\pm = M_{jk}^\pm \theta_j^{-1}$  avec :

$$M_{jk}^+ = \frac{1}{2} \begin{pmatrix} -z_j^{-1} N_{jk}^2 & N_{jk} \\ -N_{jk} & -z_j N_{jk}^2 \end{pmatrix} \quad \text{et} \quad M_{jk}^- = \frac{1}{2} \begin{pmatrix} z_j^{-1} N_{jk}^2 & N_{jk} \\ -N_{jk} & z_j N_{jk}^2 \end{pmatrix}.$$

et  $M_{jk} = M_{jk}^+ + M_{jk}^-$ .

## 7.4 Équivalences des conditions aux limites 3D

**Proposition 7** *Pour imposer faiblement la condition aux bords absorbants 3D, les deux égalités suivantes sont équivalentes pour  $j \in \mathcal{V}_j^a$  :*

$$(i) \mathbf{n}_{jk} \times \mathbf{E}_{|\Sigma_{jk}} + z_j \mathbf{n}_{jk} \times (\mathbf{n}_{jk} \times \mathbf{H}_{|\Sigma_{jk}}) = \xi[\mathbf{n}_{jk} \times \mathbf{E}^{inc} + z_j \mathbf{n}_{jk} \times (\mathbf{n}_{jk} \times \mathbf{H}^{inc})],$$

$$(ii) (M_{jk}^j)^{-\theta_j} \left( \frac{\mathbf{W}_j + \mathbf{W}_k}{2} \right) = \xi(M_{jk}^j)^{-\theta_0} \mathbf{W}^{inc}.$$

où  $\theta_0 = \begin{pmatrix} \varepsilon_0 & 0 \\ 0 & \mu_0 \end{pmatrix}$  avec  $\varepsilon_0$  et  $\mu_0$  les caractéristiques électromagnétiques du vide.

**Preuve** Il est important ici de travailler en variables conservatives. On introduit donc  $\mathbf{Q}_{E,jk} \equiv \mathbf{Q}_{E,|\Sigma_{jk}} = \varepsilon_j \mathbf{E}_{|\Sigma_{jk}}$  et  $\mathbf{Q}_{H,jk} \equiv \mathbf{Q}_{H,|\Sigma_{jk}} = \mu_j \mathbf{H}_{jk}$ . Notons que l'on a supposé  $\varepsilon_k = \varepsilon_j$  et  $\mu_k = \mu_j$ . On commence tout d'abord par ré-écrire (i) à partir de ces variables. On obtient :

$$\mathbf{n}_{jk} \times \mathbf{Q}_{E,jk} + z_j^{-1} \mathbf{n}_{jk} \times (\mathbf{n}_{jk} \times \mathbf{Q}_{H,jk}) = \xi[\mathbf{n}_{jk} \times \mathbf{Q}_E^{inc} + z_j^{-1} \mathbf{n}_{jk} \times (\mathbf{n}_{jk} \times \mathbf{Q}_H^{inc})] \quad (110)$$

On considère deux équations : la première est donnée par  $\mathbf{n}_{jk} \times (110)$  et la seconde par  $z_j \cdot (110)$ . On obtient :

$$\begin{cases} \mathbf{n}_{jk} \times [\mathbf{n}_{jk} \times \mathbf{Q}_{E,jk} + z_j^{-1} \mathbf{n}_{jk} \times (\mathbf{n}_{jk} \times \mathbf{Q}_{H,jk})] = \xi[\mathbf{n}_{jk} \times \mathbf{Q}_E^{inc} + z_j^{-1} \mathbf{n}_{jk} \times (\mathbf{n}_{jk} \times \mathbf{Q}_H^{inc})] \\ z_j (\mathbf{n}_{jk} \times \mathbf{Q}_{E,jk}) + \mathbf{n}_{jk} \times (\mathbf{n}_{jk} \times \mathbf{Q}_{H,jk}) = \xi[z_j (\mathbf{n}_{jk} \times \mathbf{Q}_E^{inc}) + \mathbf{n}_{jk} \times (\mathbf{n}_{jk} \times \mathbf{Q}_H^{inc})] \end{cases}$$

$$\Leftrightarrow \begin{cases} \mathbf{n}_{jk} \times (\mathbf{n}_{jk} \times \mathbf{Q}_{E,jk}) - z_j^{-1} \mathbf{n}_{jk} \times \mathbf{Q}_{H,jk} = \xi[\mathbf{n}_{jk} \times (\mathbf{n}_{jk} \times \mathbf{Q}_E^{inc}) - z_j^{-1} \mathbf{n}_{jk} \times \mathbf{Q}_H^{inc}] \\ z_j (\mathbf{n}_{jk} \times \mathbf{Q}_{E,jk}) + \mathbf{n}_{jk} \times (\mathbf{n}_{jk} \times \mathbf{Q}_{H,jk}) = \xi[z_j (\mathbf{n}_{jk} \times \mathbf{Q}_E^{inc}) + \mathbf{n}_{jk} \times (\mathbf{n}_{jk} \times \mathbf{Q}_H^{inc})]. \end{cases}$$

En utilisant cette fois-ci les notations matricielles, ce système donne :

$$\begin{cases} N_{jk}^2 \left( \frac{\mathbf{Q}_{E,j} + \mathbf{Q}_{E,k}}{2} \right) + z_j^{-1} N_{jk} \left( \frac{\mathbf{Q}_{H,j} + \mathbf{Q}_{H,k}}{2} \right) = \xi(N_{jk}^2 \mathbf{Q}_E^{inc} + z_j^{-1} N_{jk} \mathbf{Q}_H^{inc}) \\ -z_j N_{jk} \left( \frac{\mathbf{Q}_{E,j} + \mathbf{Q}_{E,k}}{2} \right) + N_{jk}^2 \left( \frac{\mathbf{Q}_{H,j} + \mathbf{Q}_{H,k}}{2} \right) = \xi(-z_j N_{jk} \mathbf{Q}_E^{inc} + N_{jk}^2 \mathbf{Q}_H^{inc}) \end{cases}$$

$$\Leftrightarrow \begin{pmatrix} N_{jk}^2 & z_j^{-1} N_{jk} \\ -z_j N_{jk} & N_{jk}^2 \end{pmatrix} \begin{pmatrix} \frac{\mathbf{Q}_{E,j} + \mathbf{Q}_{E,k}}{2} \\ \frac{\mathbf{Q}_{H,j} + \mathbf{Q}_{H,k}}{2} \end{pmatrix} = \xi \begin{pmatrix} N_{jk}^2 & z_j^{-1} N_{jk} \\ -z_j N_{jk} & N_{jk}^2 \end{pmatrix} \begin{pmatrix} \mathbf{Q}_E^{inc} \\ \mathbf{Q}_H^{inc} \end{pmatrix}.$$

Il se trouve que cette égalité correspond à :

$$(M_{jk}^j)^{-\theta_j} \left( \frac{\mathbf{W}_j + \mathbf{W}_k}{2} \right) = \xi(M_{jk}^j)^{-\theta_0} \mathbf{W}^{inc}.$$

■

**Proposition 8** *Pour exprimer faiblement la condition aux bords métalliques 3D, les deux égalités sont équivalentes pour  $k \in \mathcal{V}_j^m$  :*

$$(i) \mathbf{n}_{jk} \times (\mathbf{E}_{|\Sigma_{jk}}) = (\xi - 1) \mathbf{n}_{jk} \times (\mathbf{E}^{inc})$$

$$(ii) \Upsilon_{jk} \frac{\mathbf{W}_j + \mathbf{W}_k}{2} = (\xi - 1) \Upsilon_{jk} \mathbf{W}^{inc},$$

$$\text{où } \Upsilon_{jk} = \begin{pmatrix} -N_{jk} & 0_3 \\ 0_3 & 0_3 \end{pmatrix}.$$

**Preuve** On part toujours de (i) en ajoutant une égalité nulle sur le champ  $H_{|\Sigma_{jk}}$ , ce qui donne l'égalité matricielle suivant :

$$\begin{pmatrix} -N_{jk} & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \begin{pmatrix} \mathbf{E}_{|\Sigma_{jk}} \\ \mathbf{H}_{|\Sigma_{jk}} \end{pmatrix} = (\xi - 1) \begin{pmatrix} -N_{jk} & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \begin{pmatrix} \mathbf{E}^{inc} \\ \mathbf{H}^{inc} \end{pmatrix}.$$

On se sert maintenant de la moyenne des champs pour exprimer ces derniers sur l'interface  $\Sigma_{jk}$  :

$$\begin{pmatrix} -N_{jk} & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \begin{pmatrix} \frac{\mathbf{E}_j + \mathbf{E}_k}{2} \\ \frac{\mathbf{H}_j + \mathbf{H}_k}{2} \end{pmatrix} = (\xi - 1) \begin{pmatrix} -N_{jk} & 0_3 \\ 0_3 & 0_3 \end{pmatrix} \begin{pmatrix} \mathbf{E}^{inc} \\ \mathbf{H}^{inc} \end{pmatrix},$$

et aboutir à (ii) :

$$\Upsilon_{jk} \left( \frac{\mathbf{W}_j + \mathbf{W}_k}{2} \right) = (\xi - 1) \Upsilon_{jk} \mathbf{W}^{inc}.$$

■

## 7.5 Quelques propriétés des matrices antisymétriques

Ce paragraphe a pour but de fournir quelques propriétés des matrices antisymétriques, et de donner aussi certains critères permettant de prouver l'antisymétrie de matrices ayant une structure par blocs bien précise.

**Proposition 9** *Soit  $(A, B) \in (\mathbf{M}_{m \times n}(\mathbb{R}))^2$ .*

*Alors  $(\forall (X, Y) \in \mathbb{R}^{2(m \times n)}, X^t A Y = X^t B Y) \Leftrightarrow A = B$ .*

**Proposition 10** *Soit  $R \in \mathbf{M}_{m \times n}(\mathbb{R})$ .  $R$  est antisymétrique  $\Leftrightarrow (\forall X \in \mathbb{R}^{m \times n}, X^t R X = 0)$ .*

**Preuve**  $\Rightarrow$ ) Supposons  $R$  est antisymétrique et soit  $X \in \mathbb{R}^{m \times n}$ . On a que :

$$X^t R X = (R^t X)^t X = X^t (R^t X) = -X^t R X,$$

ce qui prouve le résultat.

$\Leftarrow$ ) Supposons que  $\forall \mathbf{X} \in \mathbb{R}^{m \times n} \mathbf{X}^t R \mathbf{X} = 0$ . On a :  $(\forall (\mathbf{X}, \mathbf{Y}) \in \mathbb{R}^{2(m \times n)}) :$

$$(\mathbf{X} + \mathbf{Y})^t R (\mathbf{X} + \mathbf{Y}) = 0 \Rightarrow \mathbf{X}^t (R + R^t) \mathbf{Y} + \mathbf{X}^t R \mathbf{X} + \mathbf{Y}^t R \mathbf{Y} = 0 \Rightarrow \mathbf{X}^t (R + R^t) \mathbf{Y} = 0.$$

D'après la proposition 9, on en déduit que  $R^t = -R$ , et par conséquent que  $R$  est antisymétrique.  $\blacksquare$

**Proposition 11** Soit  $M \in \mathbf{M}_{m \times n}(\mathbb{R})$  telle que  $\forall (j, k) \in \{1, \dots, n\}, j \neq k, (M_{jj}, M_{jk})$  soient respectivement antisymétrique et symétrique réelles et  $M_{kj} = -M_{jk}$ . Alors  $M$  est antisymétrique.

**Preuve** Il nous suffit de montrer d'après le lemme 10 que  $\forall \mathbf{X} \in \mathbb{R}^{m \times n}$  on ait  $\mathbf{X}^t M \mathbf{X} = 0$ . Si on pose alors  $N = M^t$  :

$$\begin{aligned} \mathbf{X}^t M \mathbf{X} &= \frac{1}{2} \mathbf{X}^t (M + N) \mathbf{X} \\ &= \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \mathbf{X}_j^t (M_{jk} + N_{jk}) \mathbf{X}_j \\ &= \frac{1}{2} \sum_{j=1}^n [\mathbf{X}_j^t (M_{jj} + N_{jj}) \mathbf{X}_j + \sum_{\substack{k=1 \\ k \neq j}}^n \mathbf{X}_k^t (M_{jk} + N_{jk}) \mathbf{X}_k]. \end{aligned}$$

Or,  $N = M^t \Rightarrow N_{jj} = M_{jj}^t$  et  $N_{jk} = M_{kj}^t$ . Mais  $M_{jj} \in \mathbf{A}_m(\mathbb{R}) \Rightarrow M_{jj}^t = -M_{jj}$  et  $M_{jk} \in \mathbf{S}_m(\mathbb{R}) \Rightarrow M_{kj}^t = M_{kj}$ . Donc :

$$\begin{aligned} \mathbf{X}^t M \mathbf{X} &= \frac{1}{2} \sum_{j=1}^n [\mathbf{X}_j^t (M_{jj} - M_{jj}) \mathbf{X}_j + \sum_{\substack{k=1 \\ k \neq j}}^n \mathbf{X}_k^t (M_{jk} + M_{kj}) \mathbf{X}_k] \\ &= \frac{1}{2} \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq j}}^n \mathbf{X}_j^t (M_{jk} + M_{kj}) \mathbf{X}_k. \end{aligned}$$

Enfin,  $M_{kj} = -M_{jk} \Rightarrow \mathbf{X}^t M \mathbf{X} = 0$ .  $\blacksquare$

**Proposition 12** Si  $A$  est antihermitienne (soit  $A^* = -A$ ) alors  $\forall \mathbf{X} \in \mathbb{C}^n, \mathbf{X}^* A \mathbf{X} \in i\mathbb{R}$ .

**Preuve** On déduit facilement que :

$$\begin{aligned}
2\Re(\mathbf{X}^* \mathbf{A} \mathbf{X}) &= \mathbf{X}^* \mathbf{A} \mathbf{X} + \overline{\mathbf{X}^* \mathbf{A} \mathbf{X}} \\
&= \mathbf{X}^* \mathbf{A} \mathbf{X} + \overline{\mathbf{X}^t \mathbf{A}^t \mathbf{X}} \\
&= \mathbf{X}^* \mathbf{A} \mathbf{X} + \mathbf{X}^* \mathbf{A}^* \mathbf{X} \\
&= \mathbf{X}^* (\mathbf{A} + \mathbf{A}^*) \mathbf{X} = 0.
\end{aligned}$$

Par conséquent,  $\mathbf{X}^* \mathbf{A} \mathbf{X} \in i\mathbb{R}$ . ■

## 7.6 Blocs matriciels du schéma GDP0

**Proposition 13** *Les blocs matriciels ont les propriétés suivantes :*

(i)  $P_{jk}^a = \begin{pmatrix} -z_j^{-1} N_{jk}^2 & 0_3 \\ 0_3 & -z_j N_{jk}^2 \end{pmatrix}$  est symétrique et positive.

(ii)  $M_{jk} = \begin{pmatrix} 0_3 & N_{jk} \\ -N_{jk} & 0_3 \end{pmatrix}$  est symétrique.

(iii)  $P_{jk}^m = \begin{pmatrix} 0_3 & N_{jk} \\ N_{jk} & 0_3 \end{pmatrix}$  est antisymétrique.

**Preuve**

(i) Avec la symétrie de  $N_{jk}^2$  (voir la proposition 6) on a :

$$(P_{jk}^a)^t = \begin{pmatrix} -z_j^{-1} (N_{jk}^2)^t & 0_3 \\ 0_3 & -z_j (N_{jk}^2)^t \end{pmatrix} = \begin{pmatrix} -z_j^{-1} N_{jk}^2 & 0_3 \\ 0_3 & -z_j N_{jk}^2 \end{pmatrix} = P_{jk}^a.$$

De plus,  $\forall X \in \mathbb{C}^{3 \times 2}$  :  $X^* P_{jk}^a X = -z_j^{-1} X_1^* N_{jk}^2 X_1 - z_j X_2^* N_{jk}^2 X_2$  et  $N_{jk}^2$  est réelle symétrique (ou encore, hermitienne) négative donc :

$$X_k^* N_{jk}^2 X_k \leq 0 \text{ pour } k \in \{1; 2\} \Rightarrow X^* P_{jk}^a X \geq 0 \text{ (car } z_j > 0).$$

(ii) La matrice  $N_{jk}$  est antisymétrique (voir la proposition 6) et donc :

$$M_{jk}^t = \begin{pmatrix} 0_3 & -N_{jk}^t \\ N_{jk}^t & 0_3 \end{pmatrix} = \begin{pmatrix} 0_3 & N_{jk} \\ -N_{jk} & 0_3 \end{pmatrix} = M_{jk}.$$

(iii) De même, nous avons aussi en raison l'antisymétrie de  $N_{jk}$  :

$$(P_{jk}^m)^t = \begin{pmatrix} 0_3 & N_{jk}^t \\ N_{jk}^t & 0_3 \end{pmatrix} = \begin{pmatrix} 0_3 & -N_{jk} \\ -N_{jk} & 0_3 \end{pmatrix} = -P_{jk}^m. \quad \text{■}$$



## 7.7 Blocs matriciels du schéma GDP1

### 7.7.1 Procédé de vectorisation

**Proposition 14** Pour  $(n, m, p, q) \in \mathbb{N}^4$ , soient  $(A, B, W) \in \mathbf{M}_{n,m}(\mathbb{C}) \times \mathbf{M}_{p,q}(\mathbb{C}) \times \mathbf{M}_{q,m}(\mathbb{C})$ . Si  $\mathbb{W} = (W_1, W_2, \dots, W_m)$  où  $(W_i)_{1 \leq i \leq m} \subset \mathbb{C}^q$ ,

$$\aleph(\mathbb{W}) = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_m \end{pmatrix} \in \mathbb{C}^{qm} \quad \text{et} \quad A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & a_{22}B & \dots & a_{2m}B \\ \vdots & \vdots & & \vdots \\ a_{n1}B & a_{n2}B & \dots & a_{nm}B \end{pmatrix} \in \mathbf{M}_{np,mq}(\mathbb{C})$$

alors :  $\aleph(B\mathbb{W}A^t) = (A \otimes B)\aleph(\mathbb{W})$

**Preuve**

$$\begin{aligned} (A \otimes B)\aleph(\mathbb{W}) &= \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & a_{22}B & \dots & a_{2m}B \\ \vdots & \vdots & & \vdots \\ a_{n1}B & a_{n2}B & \dots & a_{nm}B \end{pmatrix} \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_m \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^m a_{1j}BW_j \\ \sum_{j=1}^m a_{2j}BW_j \\ \vdots \\ \sum_{j=1}^m a_{nj}BW_j \end{pmatrix} \\ &= \aleph\left(\sum_{j=1}^m a_{1j}BW_j, \sum_{j=1}^m a_{2j}BW_j, \dots, \sum_{j=1}^m a_{nj}BW_j\right) \end{aligned}$$

De plus, étant donné que  $BW_j \in \mathbb{C}^p$ , alors pour  $i \in \llbracket 1, p \rrbracket$ ,  $(BW_j)_i = \sum_{k=1}^q b_{ik}w_{kj} = (BW)_{ij}$ .

Maintenant, prenons  $l \in \llbracket 1, n \rrbracket$ , pour chaque élément scalaire nous avons l'égalité suivante :

$$(B\mathbb{W}A^t)_{il} = \sum_{j=1}^m (BW)_{ij}a_{lj} = \sum_{j=1}^m (BW_j)_i a_{lj}$$

De ce fait, si on désigne par  $(B\mathbb{W}A^t)_l$  la  $l$ ème colonne de la matrice  $B\mathbb{W}A^t$  où  $l \in \llbracket 1, n \rrbracket$ , l'égalité scalaire précédente nous permet d'obtenir l'égalité vectorielle suivante :

$$\forall l \in \llbracket 1, n \rrbracket, (B\mathbb{W}A^t)_l = \sum_{j=1}^m BW_j a_{lj}$$

Ce qui implique :

$$\aleph\left(\sum_{j=1}^m a_{1j}BW_j, \sum_{j=1}^m a_{2j}BW_j, \dots, \sum_{j=1}^m a_{nj}BW_j\right) = \aleph((B\mathbb{W}A^t)_1, (B\mathbb{W}A^t)_2, \dots, (B\mathbb{W}A^t)_n)$$

et permet ainsi de conclure que :

$$\aleph(B\mathbb{W}A^t) = (A \otimes B)\aleph(\mathbb{W})$$

■

### 7.7.2 Propriétés du produit de Kronecker

- $(A \otimes B)^t = A^t \otimes B^t$
- $(A \otimes B)(A' \otimes B') = (AA') \otimes (BB')$
- Si  $(A, B) \in \mathbf{GL}_n(\mathbb{C}) \times \mathbf{GL}_m(\mathbb{C})$ , alors  $(A \otimes B) \in \mathbf{GL}_{np}(\mathbb{C})$  et on a :

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

- $(A, B) \in \mathbf{S}_n(\mathbb{C}) \times \mathbf{S}_m(\mathbb{C})$  ou  $(A, B) \in \mathbf{A}_n(\mathbb{C}) \times \mathbf{A}_m(\mathbb{C})$ , alors  $(A \otimes B) \in \mathbf{S}_{nm}(\mathbb{C})$
- $(A, B) \in \mathbf{S}_n(\mathbb{C}) \times \mathbf{A}_m(\mathbb{C})$  ou  $(A, B) \in \mathbf{A}_n(\mathbb{C}) \times \mathbf{S}_m(\mathbb{C})$ , alors  $(A \otimes B) \in \mathbf{A}_{nm}(\mathbb{C})$

### 7.7.3 Matrice 3D des gradients des fonctions de base P1

Il s'agit de la matrice  $P = \sum_{l=1}^3 \Phi_i^{x_l} \otimes G_{x_l}$  où :

$$P = \begin{pmatrix} 0_6 & P_{12} & P_{13} & P_{14} \\ P_{21} & 0_6 & P_{23} & P_{24} \\ P_{31} & P_{32} & 0_6 & P_{34} \\ P_{41} & P_{42} & P_{43} & 0_6 \end{pmatrix}$$

avec :

$$P_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & -\alpha_z & \alpha_y \\ 0 & 0 & 0 & \alpha_z & 0 & -\alpha_x \\ 0 & 0 & 0 & -\alpha_y & \alpha_x & 0 \\ 0 & \alpha_z & -\alpha_y & 0 & 0 & 0 \\ -\alpha_z & 0 & \alpha_x & 0 & 0 & 0 \\ \alpha_y & -\alpha_x & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$P_{13} = \begin{pmatrix} 0 & 0 & 0 & 0 & -\beta_z & \beta_y \\ 0 & 0 & 0 & \beta_z & 0 & -\beta_x \\ 0 & 0 & 0 & -\beta_y & \beta_x & 0 \\ 0 & \beta_z & -\beta_y & 0 & 0 & 0 \\ -\beta_z & 0 & \beta_x & 0 & 0 & 0 \\ \beta_y & -\beta_x & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$P_{14} = \begin{pmatrix} 0 & 0 & 0 & 0 & -\gamma_z & \gamma_y \\ 0 & 0 & 0 & \gamma_z & 0 & -\gamma_x \\ 0 & 0 & 0 & -\gamma_y & \gamma_x & 0 \\ 0 & \gamma_z & -\gamma_y & 0 & 0 & 0 \\ -\gamma_z & 0 & \gamma_x & 0 & 0 & 0 \\ \gamma_y & -\gamma_x & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$P_{21} = -P_{12}$$

$$P_{23} = \begin{pmatrix} 0 & 0 & 0 & 0 & -\delta_z & \delta_y \\ 0 & 0 & 0 & \delta_z & 0 & -\delta_x \\ 0 & 0 & 0 & -\delta_y & \delta_x & 0 \\ 0 & \delta_z & -\delta_y & 0 & 0 & 0 \\ -\delta_z & 0 & \delta_x & 0 & 0 & 0 \\ \delta_y & -\delta_x & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$P_{24} = \begin{pmatrix} 0 & 0 & 0 & 0 & -\epsilon_z & \epsilon_y \\ 0 & 0 & 0 & \epsilon_z & 0 & -\epsilon_x \\ 0 & 0 & 0 & -\epsilon_y & \epsilon_x & 0 \\ 0 & \epsilon_z & -\epsilon_y & 0 & 0 & 0 \\ -\epsilon_z & 0 & \epsilon_x & 0 & 0 & 0 \\ \epsilon_y & -\epsilon_x & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$P_{31} = -P_{13}$$

$$P_{32} = -P_{23}$$

$$P_{34} = \begin{pmatrix} 0 & 0 & 0 & 0 & -F_z & F_y \\ 0 & 0 & 0 & F_z & 0 & -F_x \\ 0 & 0 & 0 & -F_y & F_x & 0 \\ 0 & F_z & -F_y & 0 & 0 & 0 \\ -F_z & 0 & F_x & 0 & 0 & 0 \\ F_y & -F_x & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$P_{41} = -P_{14}$$

$$P_{42} = -P_{24}$$

$$P_{43} = -P_{34}$$

et :

$$\left\{ \begin{array}{lcl} \alpha_x & = & n_{1x} - n_{2x} \\ \beta_x & = & n_{1x} - n_{3x} \\ \gamma_x & = & n_{1x} - n_{4x} \\ \delta_x & = & n_{2x} - n_{3x} \\ \varepsilon_x & = & n_{2x} - n_{4x} \\ F_x & = & n_{3x} - n_{4x} \\ \alpha_y & = & n_{1y} - n_{2y} \\ \beta_y & = & n_{1y} - n_{3y} \\ \gamma_y & = & n_{1y} - n_{4y} \\ \delta_y & = & n_{2y} - n_{3y} \\ \varepsilon_y & = & n_{2y} - n_{4y} \\ F_y & = & n_{3y} - n_{4y} \\ \alpha_z & = & n_{1z} - n_{2z} \\ \beta_z & = & n_{1z} - n_{3z} \\ \gamma_z & = & n_{1z} - n_{4z} \\ \delta_z & = & n_{2z} - n_{3z} \\ \varepsilon_z & = & n_{2z} - n_{4z} \\ F_z & = & n_{3z} - n_{4z} \end{array} \right.$$

#### 7.7.4 Propriétés matricielles

**Proposition 15** *Les matrices  $\mathcal{D}$ ,  $\mathcal{Z}$  et  $\mathcal{R}$  ont les propriétés suivantes :*

- (i)  $\mathcal{D}$  est définie positive.
- (ii)  $\mathcal{Z}$  est symétrique et positive.
- (iii)  $\mathcal{R}$  est antisymétrique.

**Preuve**

(i) Tout d'abord, il est évident que  $\mathcal{D}$  est symétrique réelle.

De plus,  $\forall j \in \llbracket 1; N \rrbracket$   $\Phi_j$  et  $\theta_j$  étant symétriques, alors  $\Phi_j \otimes \theta_j$  l'est aussi compte tenu de 7.7.2 et par conséquent  $\mathcal{D}$  est symétrique diagonale par blocs.

(ii) Ensuite,  $(\Phi_1^+$  et  $Z_1)$  sont toutes deux symétriques positives, si bien que nous pouvons en déduire que  $\mathcal{Z}$  est symétrique diagonale par blocs.

(iii) Enfin,

$$- \forall j \in \llbracket 1; N-1 \rrbracket (\Phi_{j,j+1}^+ \otimes A)^t = (\Phi_{j,j+1}^+)^t \otimes A^t = \Phi_{j+1,j}^+ \otimes A.$$

Ce qui signifie donc que pour  $(j, k) \in \llbracket 1; N \rrbracket^2$ ,  $j \neq k$ ,  $\mathcal{R}_{jk} = -\mathcal{R}_{kj}$ .

-  $\forall j \in \llbracket 1; N \rrbracket$   $\Phi_j'$  est antisymétrique,  $A$  est symétrique, d'où  $\Phi_j' \otimes A$  est antisymétrique vu 7.7.2.

-  $\Phi_N^+$  est symétrique,  $\bar{Q}$  est antisymétrique, d'où  $\Phi_N^+ \otimes \bar{Q}$  est antisymétrique toujours grâce à 7.7.2, et par conséquent  $(\Phi_N' \otimes A + \Phi_N^+ \otimes \bar{Q})$  est antisymétrique par linéarité.

Ces trois points permettent d'appliquer la proposition 11 et de conclure que  $\mathcal{R}$  est antisymétrique. ■

**Proposition 16** Soit  $G = (G_{x_1}, G_{x_2}, G_{x_3})^t$  une matrice de taille  $6 \times 18$  où pour  $j = 1, 2, 3$   $G_{x_j}$  est une matrice symétrique de taille  $6 \times 6$ . Considérons que l'opérateur  $\nabla$  agit sur un vecteur  $Q \in \mathbb{C}^6$  tel que  $\nabla Q = (\partial_{x_1} Q, \partial_{x_2} Q, \partial_{x_3} Q)^t$  soit un vecteur de  $\mathbb{C}^{18}$ . Dans ces conditions et  $\forall (Q, \phi) \in \mathbb{C}^6 \times \mathbb{C}^6$  :

$$G(\bar{\phi}) \cdot \nabla Q = (\operatorname{div} G(Q)) \cdot \bar{\phi}$$

où  $\cdot$  désigne le produit scalaire canonique de  $\mathbb{R}^6$ .

**Preuve** Il est plus simple de raisonner sur  $G$  en terme de matrice plutôt qu'en terme d'opérateur. Partons du membre de gauche de l'équation. En faisant passer  $G$  de l'autre côté du produit scalaire, on obtient :

$$\begin{aligned} G(\bar{\phi}) \cdot \nabla Q &= G\bar{\phi} \cdot \nabla Q \\ &= \bar{\phi} \cdot G^t \nabla Q \end{aligned}$$

En développant maintenant le second membre et en utilisant le fait que  $G$  est un opérateur constant, on trouve :

$$\begin{aligned} G(\bar{\phi}) \cdot \nabla Q &= \bar{\phi} \cdot G^t \nabla Q \\ &= \bar{\phi} \cdot \sum_{l=1}^3 G^{x_l} \partial_{x_l} Q \\ &= \bar{\phi} \cdot \sum_{l=1}^3 \partial_{x_l} G^{x_l} Q \end{aligned}$$

Ce qui correspond directement à :

$$\begin{aligned} G(\bar{\phi}) \cdot \nabla Q &= \bar{\phi} \cdot \sum_{l=1}^3 \partial_{x_l} G^{x_l} Q \\ &= \bar{\phi} \cdot (\operatorname{div} G(Q)) \\ &= (\operatorname{div} G(Q)) \cdot \bar{\phi} \end{aligned}$$

■

## Références

- [1] D.N. Arnold. An interior penalty finite element method with discontinuous elements. *SIAM J. Numer. Anal.*, 19 :742–760, 1982.
- [2] D.N. Arnold, F. Brezzi, B. Cockburn, and L.D. Marini. Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, 39(5) :1749–1779, 2002.
- [3] F. Assous, P. Ciarlet Jr., P.A. Raviart, and E. Sonnendruker. Characterization of the singular part of the solution of Maxwell’s equations in a polyhedral domain. *Math. Meth. Appl. Sci.*, 22 :485–499, 1999.
- [4] I. Babuska and F. Ihlenburg. Finite element solution to the Helmholtz equation with high wavenumber. *Comput. Math. Appl.*, 30(9) :9–37, 1995.
- [5] P. Bonnet. Résolution des équations de maxwell stationnaires et harmoniques par une technique de volumes finis. Application à des problèmes de compatibilité électromagnétique. Technical Report NT 1998-8, ONERA, 1998.
- [6] B. Cockburn and C. Shu. TVD Runge-Kutta projection discontinuous Galerkin finite element method for conservatoin laws II : general framework. *Math. Comput.*, 52 :411–435, 1989.
- [7] G. Cohen and P. Monk. Gauss point mass lumping schemes for Maxwell’s equations. *Numer. Meth. Part. Diff. Eqns.*, 14 :63–68, 1998.
- [8] L. Demkowicz and D. Xue. Control in geometry induced in *hp* finite element simulations. *Numer. Anal. and Model.*, 2(3) :283–300, 2005.
- [9] C. Fahrat, R. Tezaur, and P. Wiedemann-Goiran. A discontinuous Galerkin method with plane waves and Lagrange multipliers for the solution of short wave exterior Helmholtz problems on unstructured meshes. *Wave motions*, 39 :307–317, 2004.
- [10] L. Fezoui, S. Lanteri, S. Lohrengel, and S. Piperno. Convergence and stability of a discontinuous Galerkin time-domain method for the 3D heterogeneous Maxwell equations on unstructured meshes. *ESAIM : Modél. Math. Anal. Numér.*, 39(6) :1149–1176, 2005.
- [11] K. Gerdes and F. Ilhenburg. On the pollution effect in FE solutions of the 3D Helmholtz equation with large wavenumber. *Comput. Methods Appl. Mech. Engrg.*, 156 :155–172, 1999.
- [12] P. Helluy. *Résolution numérique des équations de Maxwell harmoniques par une méthode d’éléments finis discontinus*. PhD thesis, Ecole Nationale Supérieure de l’Aéronautique, 1994.
- [13] P. Helluy and S. Dayma. Convergence d’une approximation discontinus des systèmes du premier ordre. *C. R. Acad. Sci. Paris*, 319 :1331–1335, 1994.
- [14] P. Helluy, P. Mazet, and P. Klotz. Sur une approximation en domaine non borné des équations de Maxwell stationnaires : comportement asymptotique. *La Recherche Aérospatiale*, 5 :365–377, 1994.
- [15] J. Hesthaven and T. Warburton. Nodal high-order methods for symmetric hyperbolic equations. *SIAM J. Numer. Anal.*, 36 :935–952, 1999.

- [16] J. Hesthaven and T. Warburton. Discontinuous Galerkin methods for the time-domain Maxwell's equations : an introduction. *ACES Newsletter*, 19 :10–29, 2004.
- [17] J. Hesthaven and T. Warburton. High-order nodal Discontinuous Galerkin methods for the Maxwell eigenvalue problem. *Phil. Trans. Roy. Soc. London A*, 362 :493–524, 2004.
- [18] J. Hesthaven and T. Warburton. The role of the penalty in the local discontinuous Galerkin method for Maxwell's eigenvalue problem. *Comput. Meth. App. Mech. Engng.*, 2006. to appear.
- [19] T. Hill and W. Reed. Triangular mesh methods for the neutron transport equation. Technical Report LA-UR-73-479, Los Alamos Scientific Laboratory, 1973.
- [20] R. Hiptmair. Higher order Whitney forms. *Prog.In Electr. Res. (PIER)*, 32 :271–299, 2001.
- [21] P. Houston, I. Perugia, and D. Schötzau. Mixed discontinuous Galerkin approximation of the Maxwell operator. *SIAM J. Numer. Anal.*, 42 :434–459, 2004.
- [22] F. Ihlenburg, editor. *Finite element analysis of acoustic scattering*. Springer-Verlag, Berlin, 1998.
- [23] J. Douglas Jr. and T. Dupont. Interior penalty procedures for elliptic and parabolic Galerkin methods. In *Lecture notes in Physics*, volume 58. Springer-Verlag Berlin, 1976.
- [24] S. Lanteri. Parallel solutions of compressible flows using overlapping and non-overlapping mesh partitioning strategies. *Parallel Comput.*, 22 :943–968, 1996.
- [25] P. Monk, editor. *Finite element methods for Maxwell's equations*. Oxford University Press, 2003.
- [26] P. Monk and G. R. Richter. A discontinuous Galerkin method for linear symmetric hyperbolic systems in inhomogeneous media. *J. of Scient. Comput.*, 22 & 23 :443–477, 2005.
- [27] J.C. Nédélec. Mixed finite elements in  $\mathbb{R}^3$ . *Numer. Math.*, 35 :315–341, 1980.
- [28] J.C. Nédélec. A new family of mixed finite elements in  $\mathbb{R}^3$ . *Numer. Math.*, 50 :97–112, 1986.
- [29] I. Perugia and D. Schötzau. The *hp*-local discontinuous Galerkin method for low-frequency time-harmonic Maxwell equations. *Math. Comp.*, 72 :1179–1214, 2003.
- [30] I. Perugia, D. Schötzau, and P. Monk. Stabilized interior penalty method for low-frequency time-harmonic Maxwell equations. *Comput. Meth. Appl. Mech. Engrg.*, 191 :4675–4697, 2002.
- [31] S. Piperno.  $l^2$ -stability of the upwind first order finite volume scheme for the Maxwell equations in two and three dimensions on arbitrary unstructured meshes. *RAIRO : Modél. Math. Anal. Numér.*, 34 :139–158, 2000.
- [32] S. Piperno and L. Fezoui. A centered discontinuous Galerkin finite volume scheme for the 3D heterogeneous Maxwell equations on unstructured meshes. Technical Report 4733, INRIA, 2003.



- [33] S. Piperno, M. Remaki, and L. Fezoui. A nondiffusive finite volume scheme for the three-dimensional Maxwell's equations on unstructured meshes. *SIAM J. Num. Anal.*, 39(6) :2089–2108, 2002.
- [34] J. Rauch. Asymptotic behaviour of solutions to hyperbolic partial differential equations with zero speeds. *Comm. Pure Appl. Math.*, 31 :431–480, 1978.
- [35] M. Remaki. A new finite volume scheme for solving Maxwell's system. *COMPEL*, 19 :913–931, 2000.
- [36] G.L.G. Sleijpen and D.R. Fokkema. BiCGstab( $\ell$ ) for linear equations involving unsymmetric matrices with complex spectrum. *Electron. Trans. Numer. Anal.*, 1 :11–32 (electronic only), 1993.
- [37] J. Steger and R.F. Warming. Flux vector splitting for the inviscid gas dynamic with applications to finite difference methods. *J. of Comp. Phys.*, 40 :263–293, 1981.
- [38] M.F. Wheeler. An elliptic collocation-finite element method with interior penalties. *SIAM J. Numer. Anal.*, 15 :152–161, 1978.
- [39] K.S. Yee. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Trans. Antennas and Propagation*, AP-16 :302–307, 1966.



---

Unité de recherche INRIA Sophia Antipolis  
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes  
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique  
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

---

Éditeur  
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)  
<http://www.inria.fr>  
ISSN 0249-6399